NATIONAL TECHNICAL UNIVERSITY OF ATHENS

INTERDEPARTMENTAL PROGRAM

OF POSTGRADUATE STUDIES

«WATER RESOURCES SCIENCE AND TECHNOLOGY»

Refining the working hypotheses of parameter identification in hydrological modelling: the concept of stochastic calibration

M.Sc. Thesis

Vasileios Kourakos

Supervisor: **Andreas Efstratiadis, Assistant Professor, NTUA**

«WATER RESOURCES SCIENCE AND TECHNOLOGY»

Athens, November 2021

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

INTERDEPARTMENTAL PROGRAM OF POSTGRADUATE STUDIES

«WATER RESOURCES SCIENCE AND TECHNOLOGY»

Refining the working hypotheses of parameter identification in hydrological modelling: the concept of stochastic calibration

M.Sc. Thesis

Vasileios Kourakos

Supervisor: **Andreas Efstratiadis, Assistant Professor, NTUA**

Athens, November 2021

ACKNOWLEDGEMENTS (IN GREEK)

Με την ολοκλήρωση των μεταπτυχιακών μου σπουδών, φτάνει ένα ακόμη κεφάλαιο στο τέλος του. Μία περίοδος που συνοδεύτηκε παρά τις όποιες δυσκολίες από πολύ ενθουσιασμό και χαρά και από την οποία αποχωρώ γεμάτος όμορφες αναμνήσεις. Για την δημιουργία όμως των παραπάνω ευχάριστων συνθηκών συνέβαλαν αρκετοί άνθρωποι τους οποίους θα ήθελα να ευχαριστήσω έστω κι από εδώ.

Αρχικά, θα ήθελα να ευχαριστήσω τον Δρ. Πολιτικό Μηχανικό κ. Ανδρέα Ευστρατιάδη, επίκουρο καθηγητή ΕΜΠ. Μέσα από τα μαθήματα του μεταπτυχιακού κατά το 2ο εξάμηνο, μου δόθηκε η ευκαιρία να γνωρίσω έναν άνθρωπο με μεγάλη όρεξη να μεταδώσει τις γνώσεις του πάνω στο αντικείμενο του Υδρολόγου και του Υδραυλικού Μηχανικού. Θα ήθελα λοιπόν να τον ευχαριστήσω θερμά, για την εμπιστοσύνη και τη σιγουριά που έδειξε στο πρόσωπό μου, αναθέτοντας μου ένα πολύ ενδιαφέρον και πρωτότυπο θέμα, το οποίο έχει αναπτύξει ο ίδιος και η ομάδα του. Ακόμη, τον ευχαριστώ για την υπομονή του, ειδικά στην περίπτωση των μακροσκελών μου e-mails (το περιεχόμενο των οποίων μετά από κάποιο διάστημα μετά βίας κατανοώ κι εγώ ο ίδιος), καθώς και για την διαρκή καθοδήγηση του σε ό,τι χρειαζόμουν.

Φυσικά, δεν θα μπορούσα να παραλείψω και τον Δρ. Πολιτικό Μηχανικό και συνεπιβλέπον στην διπλωματική μου, Γιάννη Τσουκαλά. Η Υδροπληροφορική ήταν το μάθημα που συνδύασε δύο γνωστικά πεδία που χαίρομαι να απασχολούν τον νου μου, την Υδρολογία και τον Προγραμματισμό. Έτσι, βλέποντας με πόση ευκολία και ευχάριστη διάθεση προσέγγιζε ο Γιάννης την υλοποίηση προβλημάτων υδατικών πόρων σε προγραμματιστικό περιβάλλον, δεν μπορούσε παρά να προκαλέσει τον θαυμασμό μου και να έχει την αμέριστη προσοχή μου κατά τις παραδόσεις του μαθήματος. Τον ευχαριστώ ιδιαίτερα που μου έδωσε την ευκαιρία να κάνω τη μετάβαση σε μια νέα γλώσσα προγραμματισμού στα πλαίσια της διπλωματικής μου εργασίας, όπως και για τα χρήσιμα tips κατά την εκμάθηση της R. Επίσης, τον ευχαριστώ για την πολύτιμη και άμεση βοήθεια που μου προσέφερε, κάθε φορά που την χρειαζόμουν.

Εύχομαι οι εξ αποστάσεως συνθήκες που επιβάλλονταν κατά την εκπόνηση της διπλωματικής να ήταν διαφορετικές, ώστε να είχα την ευκαιρία να μάθω ακόμη περισσότερα κοντά τους.

Θα ήθελα επίσης να ευχαριστήσω τους διδάκτορες Παναγιώτη Δημητριάδη, Άνυ Ηλιοπούλου, Διονύση Νικολόπουλο και τη Γεωργία Παπαδονικολάκη, για τη βοήθεια που μου πρόσφεραν κατά την διάρκεια του μεταπτυχιακού, τις ενδιαφέρουσες συζητήσεις, καθώς και την περισσή όρεξη να μεταδώσουν τις γνώσεις τους, πράγματα τα οποία με

έκαναν να αγαπήσω ακόμη περισσότερο τα μαθήματα τους. Μαζί με τον Γιάννη, αποτελούν όλοι τους λαμπρό παράδειγμα νέων επιστημόνων.

Φυσικά, η καθημερινότητα της ακαδημαϊκής ζωής δεν θα ήταν η ίδια χωρίς τη συμβολή των συμφοιτητών μου Ευσταθία Λάσκαρη, Νίκο Κούρμπαση, Κωστή Μπριντάκη και Πάνο Σουσούρη. Τους ευχαριστώ πολύ για τις στιγμές γέλιου που μου χάρισαν και για τις ενδιαφέρουσες συζητήσεις που έχω να θυμάμαι μαζί τους.

Το τελευταίο και μεγαλύτερο «ευχαριστώ» ανήκει στην οικογένεια μου, τους γονείς μου, Γιώτα και Γιάννη, και τα αδέρφια μου, Εύα, Δημοσθένη και Ηλιάνα. Δεν έπαψαν ποτέ να είναι δίπλα μου, για να με στηρίξουν σε κάθε μου δυσκολία και σε ό,τι επιλέξω.

Βασίλειος Κουράκος

Αθήνα, Νοέμβριος 2021

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The aim of this study is the introduction of a promising strategy for hydrological calibration, which utilizes synthetic data as drivers to identify model parameters and evaluates the adjusted model structure against the full historical sample. These synthetic time series incorporate the marginal properties and dependence structure of the observed data across multiple time scales.

One of the main advantages of this methodology against conventional split-sample approaches is the estimation of more robust and "stable" parameter sets. This is due to the model being trained over a much longer dataset and extended hydroclimatic conditions. Another important advantage of this methodology is that the defined modelling scheme is evaluated against the full set of the observed data, hence the validation data set is also extended.

In order to prove that the proposed calibration framework is independent of the chosen model, five lumped hydrological models of varying complexity were used for the testing of the calibration scheme. Initially, a proof-of-concept was employed on a representative catchment and across two time scales, monthly and daily, by using for each scale analysis two hydrological models with alternative parameterization. The proposed calibration technique performed equally well as the classical split-sample scheme at the monthly time scale, whereas it demonstrated slightly lower performance at the daily scale.

After proving the functionality of the stochastic calibration for this case study, this strategy was tested against a large set of catchments of the MOPEX database, at the monthly scale, to further reinforce the validity of the recommended methodology. From this large scale analysis it is deduced that the stochastic calibration outperformed the split-sample approach in more than half of the examined cases, regardless of the chosen hydrological model. In addition, stochastic calibration proved to be independent of the model structure's complexity.

Keywords: Hydroinformatics, Hydrological Models, Stochastics, Stochastic Calibration

EXTENDED ABSTRACT (IN GREEK)

Ξεκαθαρίζοντας τις υποθέσεις εργασίας στον προσδιορισμό των παραμέτρων των υδρολογικών μοντέλων: η έννοια της στοχαστικής βαθμονόμησης

Εισαγωγή

Στη βιβλιογραφία απαντώνται διάφορες μέθοδοι βαθμονόμησης υδρολογικών μοντέλων με πιο κλασσική αυτή του Vit Klemeš, κατά την οποία το ιστορικό δείγμα διαχωρίζεται σε δύο επιμέρους σετ δεδομένων. Αναλυτικότερα, το πρώτο σετ (περίοδος βαθμονόμησης) στόχο έχει τη ρύθμιση του υδρολογικού μοντέλου στις υδρομετεωρολογικές συνθήκες που επικρατούν κατά την περίοδο αυτή, με την κατάλληλη προσαρμογή των παραμέτρων του, ενώ το δεύτερο σετ (περίοδος επαλήθευσης) εξυπηρετεί στον έλεγχο της αναπαραγωγής της παρατηρημένης απόκρισης του υδρολογικού συστήματος κατά τη χρονική αυτή περίοδο.

Έκτοτε, ένας σημαντικός αριθμός επιστημονικών ερευνών έχει διεξαχθεί σε μια προσπάθεια βελτίωσης της μεθόδου, αλλά και αντιμετώπισης διαφόρων μειονεκτημάτων που ενέχει η χρήση της μεθόδου, οδηγώντας έτσι σε διάφορες παραλλαγές της.

Όσον αφορά τα μειονεκτήματα της μεθόδου διαχωρισμού του ιστορικού σετ δεδομένων σε περιόδους βαθμονόμησης και επαλήθευσης, τέτοια μπορούν να θεωρηθούν τα ακόλουθα:

- Το δείγμα της επαλήθευσης εμπεριέχει πολύτιμη υδρολογική πληροφορία, η οποία πάντα πρέπει να θυσιαστεί για τους σκοπούς της επαλήθευσης του βαθμονομημένου μοντέλου.
- Ακόμη, λόγω πιθανής ασυμφωνίας μεταξύ της απόδοσης του ρυθμισμένου μοντέλου κατά την περίοδο βαθμονόμησης και επαλήθευσης, υπάρχει ανάγκη για επιμήκυνση του χρονικού ορίζοντα της ρύθμισης, ώστε το μοντέλο να είναι ακριβέστερα βαθμονομημένο στην υδρολογική δίαιτα της περιοχής μελέτης. Ωστόσο, αυτό δεν είναι πάντα εφικτό, αφού τα ιστορικά δεδομένα δεν είναι πάντα αντιπροσωπευτικά του υδρολογικού καθεστώτος που κυριαρχεί στη περιοχή, ειδικά σε περιπτώσεις που το μήκος των παρατηρήσεων είναι μικρό.
- Επιπλέον, η εφαρμογή της μεθόδου απαιτεί το ίδιο μήκος των παρατηρημένων χρονοσειρών για τις υδρολογικές διεργασίες εισόδου και εξόδου στο μοντέλο. Ως εκ τούτου, οποιαδήποτε περίοδος δεδομένων δεν ανήκει στο κοινό διάστημα μετρήσεων

μεταξύ των μεγεθών εισόδου και εξόδου, θα πρέπει να αγνοηθεί. Αυτό το πρόβλημα αντιμετωπίζεται συχνά κατά την υδρολογική μοντελοποίηση, αφού τα δείγματα μετρήσεων βροχής συνήθως ξεπερνούν εκείνα της απορροής.

Μεθοδολογική προσέγγιση της μεθοδολογίας της στοχαστικής βαθμονόμησης

Στα πλαίσια της παρούσας μεταπτυχιακής διπλωματικής εργασίας παρουσιάζεται και εφαρμόζεται μια νέα μεθοδολογία βαθμονόμησης υδρολογικών μοντέλων, αυτή της στοχαστικής βαθμονόμησης.

Η καινοτόμα αυτή τεχνική βαθμονόμησης επινοήθηκε από τους Α. Ευστρατιάδης, Ι. Τσουκαλάς και Π. Κοσσιέρης, με σκοπό να αντιμετωπιστούν τα θέματα της κλασσικής μεθόδου βαθμονόμησης που απαριθμήθηκαν παραπάνω. Συγκεκριμένα, η μεθοδολογία αυτή χρησιμοποιεί συνθετικά δεδομένα εισόδου και εξόδου, μεγάλου μήκους (εκατοντάδων ή χιλιάδων ετών) για τη διαδικασία της βαθμονόμησης, ενώ το ρυθμισμένο μοντέλο επαληθεύεται για το σύνολο της περιόδου των ιστορικών παρατηρήσεων. Οι συνθετικές χρονοσειρές παράγονται με χρήση ενός στοχαστικού μοντέλου, το οποίο θα πρέπει να εξασφαλίζει την αναπαραγωγή όλης της στατιστικής πληροφορίας που περιγράφει το πλήρες υδρολογικό καθεστώς που επικρατεί στην περιοχή μελέτης. Η πληροφορία αυτή εμπεριέχεται στις ιστορικές χρονοσειρές, αλλά είναι δύσκολα ανιχνεύσιμη, ειδικά στην περίπτωση χρονοσειρών μικρού μήκους, όπου η δυσκολία εντείνεται λόγω του ότι ένα τμήμα των δεδομένων αυτών δεσμεύεται για τους σκοπούς της επαλήθευσης του μοντέλου.

Κατά συνέπεια, τα μειονεκτήματα που πηγάζουν κατά την εφαρμογή της μεθόδου διαχωρισμού του ιστορικού δείγματος εξαλείφονται, εφόσον τα σημαντικά μεγαλύτερου μήκους συνθετικά δεδομένα που διατίθενται για τους σκοπούς της ρύθμισης, οδηγούν τελικά σε πιο «εύρωστα» σετ παραμέτρων και περισσότερο σταθερή ικανότητα πρόβλεψης του μοντέλου. Επιπροσθέτως, η προσαρμοσμένη στην περιοχή μελέτης δομή του μοντέλου επαληθεύεται πια με χρήση όλου του ιστορικού δείγματος, με αποτέλεσμα να επιμηκύνεται και η περίοδος επαλήθευσης. Δεδομένων των συνθηκών αυτών, το δίλλημα σχετικά με το ποιο τμήμα των δεδομένων θα δεσμευτεί για τη διαδικασία της βαθμονόμησης και ποιο για την επαλήθευση, δεν υφίσταται πια.

Όσον αφορά τη παραγωγή των στοχαστικών χρονοσειρών, το υπεύθυνο μοντέλο θα πρέπει να τηρεί τις εξής προδιαγραφές:

- Περιεκτική αναπαράσταση και προσομοίωση των υδρομετεωρολογικών δεδομένων εισόδου και εξόδου, αναπαράγοντας όλα τα στατιστικά χαρακτηριστικά των ιστορικών παρατηρήσεων, με ανάθεση κατάλληλου μοντέλου ανάλογα με το είδος της διεργασίας προς προσομοίωση και της προς εξέταση χρονικής κλίμακας. Αυτό

επιτυγχάνεται με χρήση σχημάτων στοχαστικής προσομοίωσης που λαμβάνουν υπόψη τις κύριες ιδιαιτερότητες των υδρομετεωρολογικών διεργασιών, ήτοι μη-γκαουσιανή στατιστική συμπεριφορά, περιοδικότητα, διαλείπουσα φύση και μη αντιστρεψιμότητα. Ακόμη, σε περίπτωση ακραίων γεγονότων (π.χ., πλημμυρών), θα πρέπει να δοθεί ιδιαίτερη έμφαση στη μορφή της ουράς, η οποία καθορίζει το μέγεθος και τη συχνότητα των ακραίων γεγονότων.

- Αναπαράσταση των αυτοσυσχετίσεων (βραχυπρόθεσμων και μακροπρόθεσμων) και των ετεροσυσχετίσεων, οι οποίες αποτελούν στατιστικά μέτρα για τις σχέσεις αίτιου και αιτιατού μεταξύ των διεργασιών του υδρολογικού κύκλου.
- Στατιστική συνέπεια σε πολλαπλές χρονικές κλίμακες, ώστε να διασφαλισθεί η αναπαραγωγή της πιθανοτικής και στοχαστικής συμπεριφοράς των διεργασιών προς μοντελοποίηση όχι μόνο στην κλίμακα της υδρολογικής προσομοίωσης (π.χ., ημερήσια ή μηνιαία), αλλά και σε μεγαλύτερες χρονικά κλίμακες (π.χ., ετήσια).

Υδρολογικά μοντέλα μελέτης

Προκειμένου να εξεταστεί η λειτουργικότητα του προτεινόμενου πλαισίου βαθμονόμησης, πραγματοποιήθηκαν αναλύσεις με χρήση ενός αριθμού υδρολογικών μοντέλων μεταβλητής πολυπλοκότητας, τα οποία είτε αναπτύχθηκαν σε περιβάλλον της γλώσσα προγραμματισμού R για τους σκοπούς της εργασίας, είτε έχουν ήδη προγραμματιστεί σε γλώσσα R.

Καθένα από αυτά τα μοντέλα είναι ένα αδιαμέριστο εννοιολογικό προσομοίωμα, το οποίο κάνει χρήση ενός σετ μαθηματικών εξισώσεων για την περιγραφή των κύριων υδρολογικών μηχανισμών που λαμβάνουν χώρα σε κλίμακα υδρολογικής λεκάνης απορροής. Ως δεδομένα εισόδου το κάθε μοντέλο δέχεται τη μέση βροχόπτωση και τη δυνητική εξατμισοδιαπνοή, σε χρονική κλίμακα ημέρας ή μήνα, ανάλογα με το υδρολογικό μοντέλο.

Τα μοντέλα αυτά είναι τα εξής:

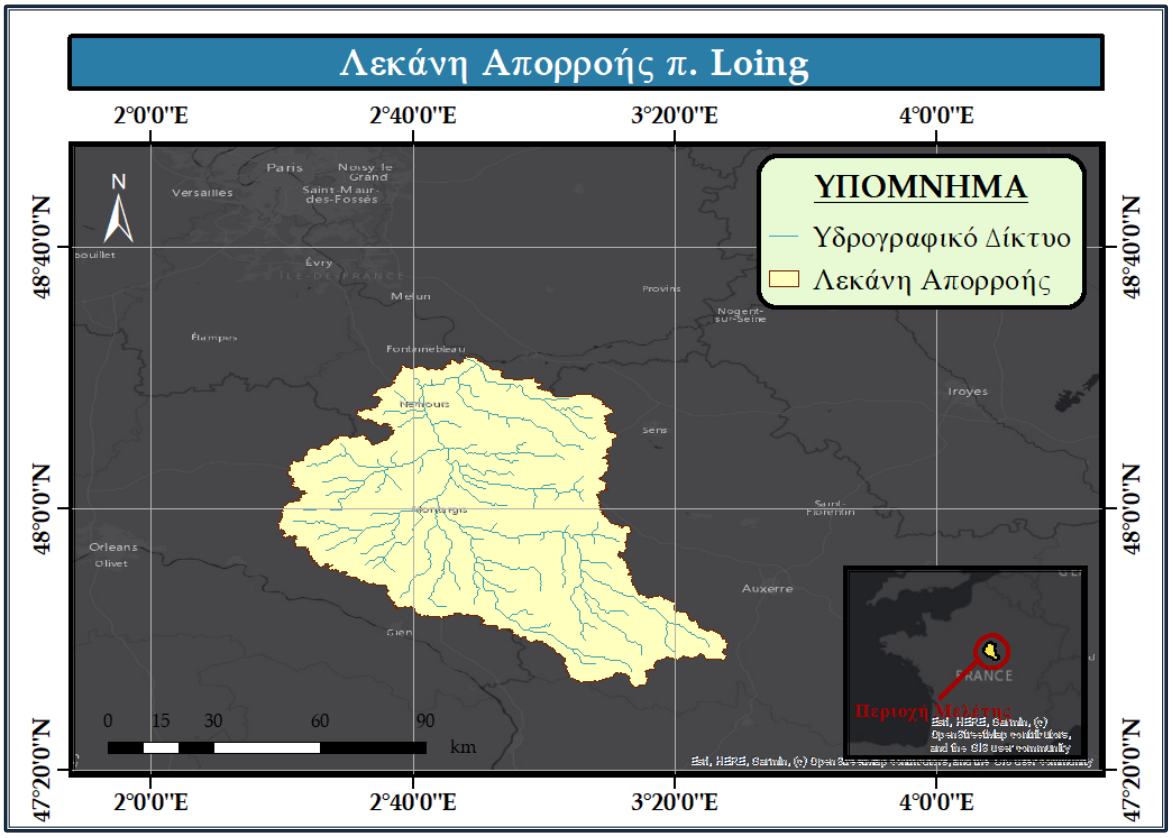
- Zygos4P: ένα μοντέλο υδατικού ισοζυγίου τεσσάρων παραμέτρων, που έχει βασιστεί σε μια απλοποιημένη εκδοχή του υδρολογικού μοντέλου «Ζυγός»
- Zygos6P: το μοντέλο αυτό χαρακτηρίζεται από έξι παραμέτρους και βασίστηκε στη δομή του προηγούμενου μοντέλου, εφαρμόζοντας μερικές τροποποιήσεις
- GR2M: το μοντέλο αυτό ανήκει στην οικογένεια των μοντέλων GR, ένα σετ εννοιολογικών αδρομερών υδρολογικών μοντέλων που αναπτύχθηκαν για συγκεκριμένο βήμα προσομοίωσης το καθένα. Το συγκεκριμένο μοντέλο έχει αναπτυχθεί για μηνιαίο βήμα προσομοίωσης και χρησιμοποιεί δύο παραμέτρους.

- GR4J: ακόμη ένα μοντέλο που ανήκει στην ομάδα των μοντέλων GR, το οποίο είναι ένα ημερήσιο υδρολογικό μοντέλο τεσσάρων παραμέτρων
- GR6J: ένα ακόμη ημερήσιο υδρολογικό μοντέλο της σειράς GR, που αποτελείται από έξι παραμέτρους και προέκυψε μετά από διαρκείς τροποποιήσεις εκείνου των τεσσάρων παραμέτρων

Ως μέτρο αξιολόγησης της καλής προσαρμογής του μοντέλου χρησιμοποιήθηκε ο δείκτης απόδοσης Nash-Sutcliffe (NSE). Όσον αφορά τον αλγόριθμο βελτιστοποίησης για τη διαδικασία της ρύθμισης, για τα μοντέλα της κατηγορίας GR έγινε χρήση ενσωματωμένου αλγορίθμου βελτιστοποίησης στον κώδικα των μοντέλων, ο οποίος συνδυάζει τοπική και ολική προσέγγιση αναζήτησης του βέλτιστου σετ τιμών των παραμέτρων, ενώ για τα μοντέλα της κατηγορίας Zygos έγινε χρήση του εξελικτικού αλγορίθμου ανόπτησης-απλόκου, ενός υβριδικού σχήματος που συνδυάζει τα πλεονεκτήματα των μεθόδων τοπικής και ολικής αναζήτησης.

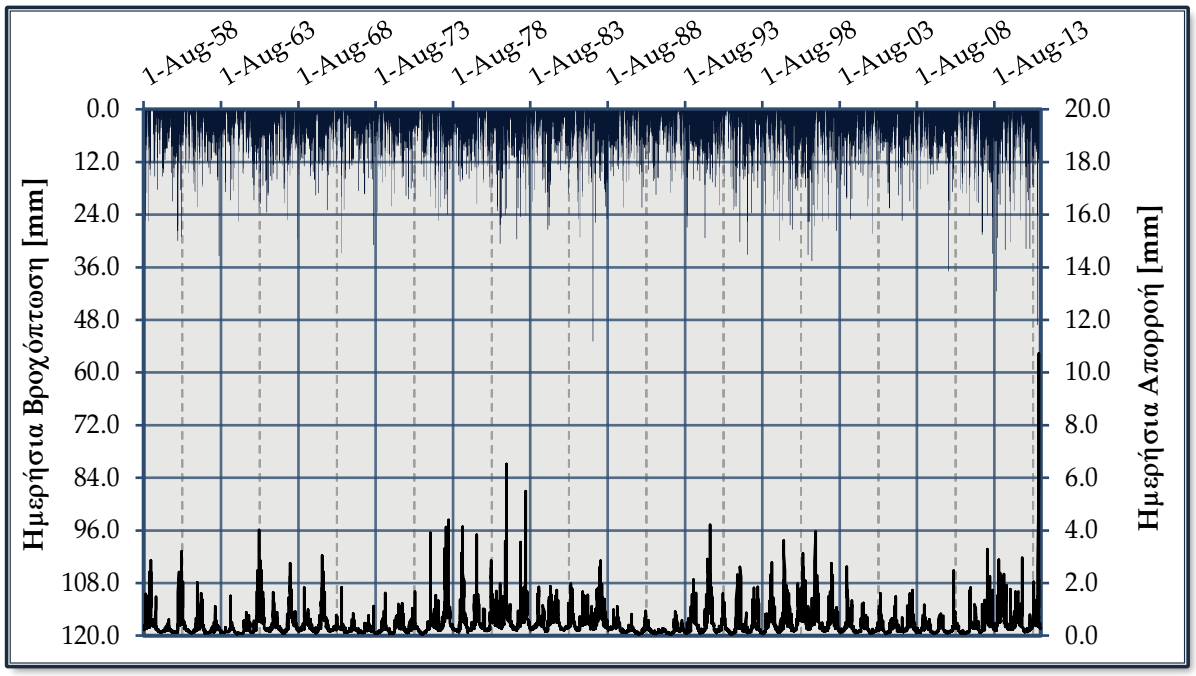
Εφαρμογή μεθοδολογίας σε περιοχή μελέτης

Προκειμένου να εξεταστεί η ισχύς της προτεινόμενης μεθόδου, επιλέχθηκε μια συγκεκριμένη λεκάνη απορροής, αυτή του ποταμού Loing (**Σχήμα 1**), έκτασης 3900 km² και μέσου τοπογραφικού υψομέτρου 148 m.



Σχήμα 1 | Λεκάνη απορροής π. Loing και υδρογραφικό δίκτυο

Για την λεκάνη αυτή είναι διαθέσιμα δεδομένα 58 ετών (από 01/08/1958 έως 31/07/2016) σε ημερήσια κλίμακα (**Σχήμα 2**), συνεπώς επιλέγεται η ανάλυση να πραγματοποιηθεί για δύο σενάρια: ένα όπου η μεθοδολογία εφαρμόζεται σε μηνιαία κλίμακα, συναθροίζοντας τα ημερήσια δεδομένα σε κλίμακα μήνα, κι ένα σε ημερήσια κλίμακα. Για την ανάλυση σε μηνιαία κλίμακα έγινε χρήση των μοντέλων Zygos6P και GR2M, ενώ για αυτή σε ημερήσια κλίμακα εξετάστηκαν τα μοντέλα GR4J και GR6J.

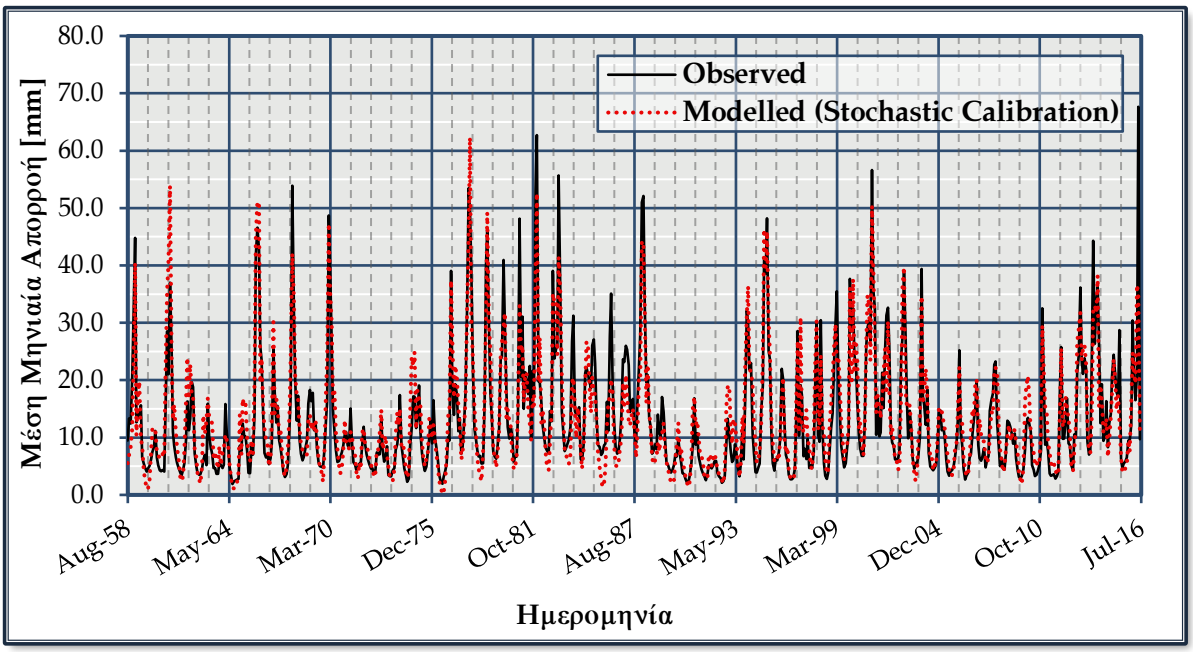


Σχήμα 2 | Ιστορικές χρονοσειρές βροχής και απορροής σε ημερήσια κλίμακα (1η Αυγούστου 1958 έως 31η Ιουλίου 2016)

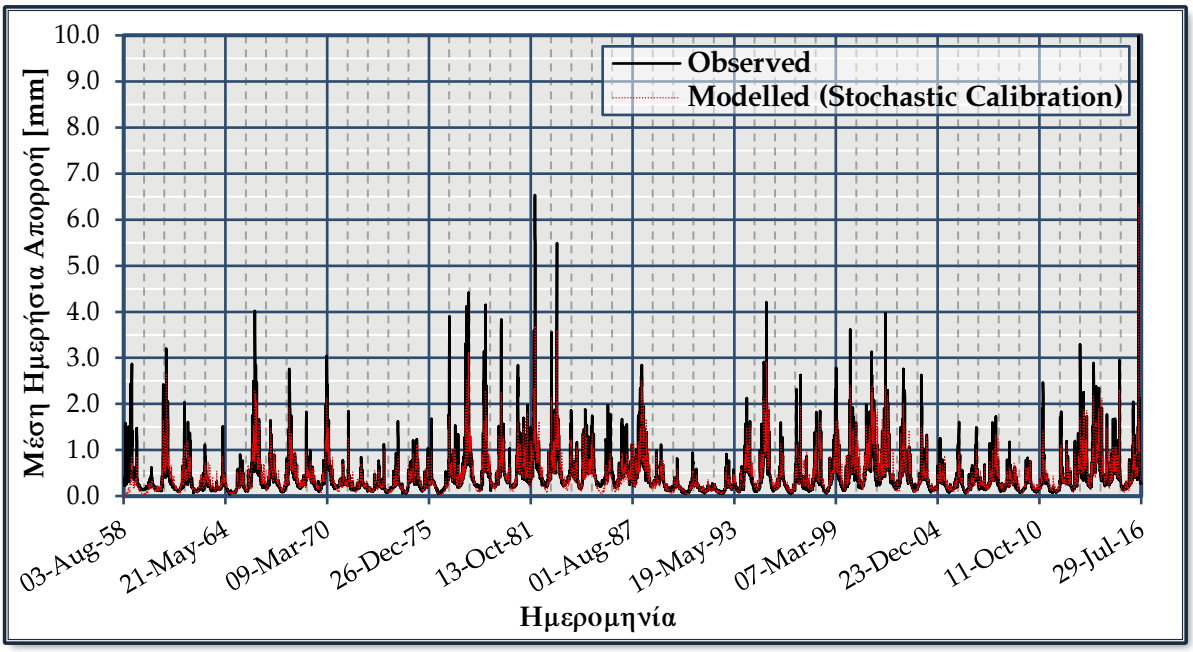
Σε κάθε ανάλυση αρχικά πραγματοποιούταν η προσαρμογή κάθε υδρολογικού μοντέλου με χρήση της μεθοδολογίας διαχωρισμού του ιστορικού δείγματος σε δύο ίσες περιόδους, η πρώτη για βαθμονόμηση και η δεύτερη για επαλήθευση. Έπειτα, ακολουθούσε η στοχαστική βαθμονόμηση του κάθε υδρολογικού μοντέλου, για την οποία οι συνθετικές χρονοσειρές βροχής και απορροής παράχθηκαν με χρήση του στοχαστικού μοντέλου *anySim*, ενώ αυτή της δυνητικής εξατμισοδιαπνοής εκτιμήθηκε με βάση τις μέσες μηνιαίες και ημερήσιες τιμές που προέκυψαν από την αντίστοιχη ιστορική χρονοσειρά.

Στο **Σχήμα 3** παρουσιάζεται ένα ενδεικτικό παράδειγμα της ανάλυσης σε μηνιαία κλίμακα (ανάλυση με το υδρολογικό μοντέλο Zygos6P), όπου ο συντελεστής NSE προέκυψε ίσος με 0.84 για χρήση και των δύο τεχνικών ρύθμισης, ενώ στο **Σχήμα 4** παρουσιάζεται ένα ενδεικτικό παράδειγμα της ανάλυσης σε επίπεδο ημερήσιας κλίμακας (ανάλυση με το υδρολογικό μοντέλο GR4J), για το οποίο προέκυψε συντελεστής NSE ίσος με 0.81 για την περίπτωση της στοχαστική βαθμονόμησης και 0.90 για την περίπτωση

της βαθμονόμησης με ιστορικά δεδομένα (η τιμή αυτή αφορά μόνο την περίοδο επαλήθευσης κι όχι όλο το ιστορικό δείγμα).



Σχήμα 3 | Οπτική επισκόπηση της συμφωνίας μεταξύ των παρατηρημένων (observed) και μοντελοποιημένων (modelled) τιμών απορροής με χρήση της μεθοδολογίας της στοχαστικής βαθμονόμησης και του υδρολογικού μοντέλου Zygos6P



Σχήμα 4 | Οπτική επισκόπηση της συμφωνίας μεταξύ των παρατηρημένων (observed) και μοντελοποιημένων (modelled) τιμών απορροής με χρήση της μεθοδολογίας της στοχαστικής βαθμονόμησης και του υδρολογικού μοντέλου GR4J

Συμπερασματικά, τα αποτελέσματα της ανάλυσης σε μηνιαία κλίμακα για τα δύο υδρολογικά μοντέλα που εξετάστηκαν, αποδεικνύουν ότι το προτεινόμενο μεθοδολογικό πλαίσιο ρύθμισης με συνθετικά δεδομένα λειτουργεί εξίσου καλά με το κλασσικό σχήμα βαθμονόμησης που απαιτεί τον διαχωρισμό των ιστορικών χρονοσειρών σε δεδομένα βαθμονόμησης και επαλήθευσης. Επιπλέον, οι ρυθμισμένες τιμές των παραμέτρων των μοντέλων βρίσκονται σε συμφωνία.

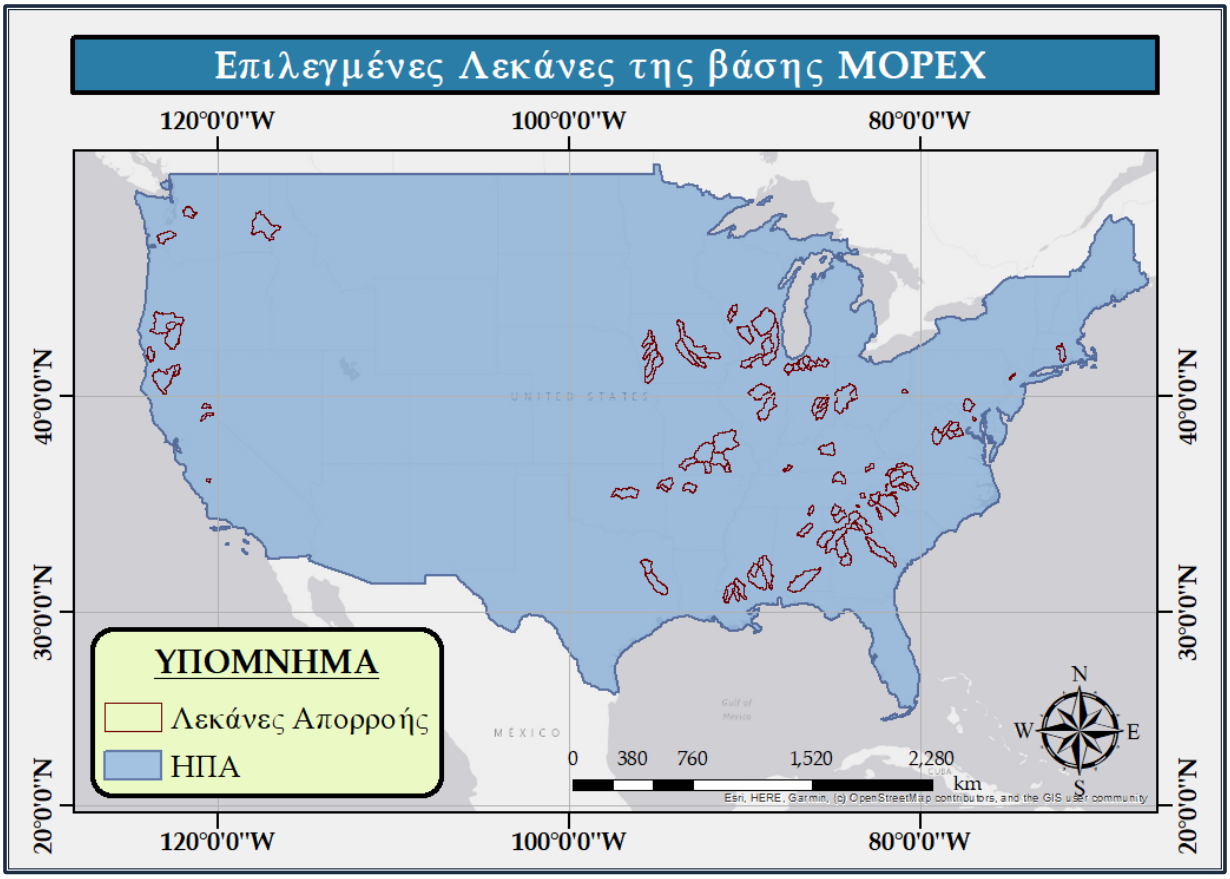
Στην περίπτωση της ανάλυσης σε ημερήσια κλίμακα, τα αποτελέσματα για τα δύο υδρολογικά μοντέλα υποδεικνύουν ότι το κλασσικό σχήμα αποδίδει ελαφρώς καλύτερα από αυτό της στοχαστικής βαθμονόμησης, ωστόσο χωρίς η απόκλιση αυτή να ακυρώνει τις προοπτικές της μεθόδου σε αυτή τη χρονική κλίμακα. Όμως, θα πρέπει να σημειωθεί ότι και για τα δύο εξεταζόμενα μοντέλα ημερήσιας κλίμακας υπήρξε μια αισθητή απόκλιση μεταξύ των τιμών των παραμέτρων που προέκυψαν από κάθε μέθοδο ρύθμισης.

Επίσης, τα αποτελέσματα αυτής της αρχικής διερεύνησης αποτελούν μια πρώτη ένδειξη ότι η εφαρμογή του μεθοδολογικού πλαισίου της στοχαστικής βαθμονόμησης είναι ανεξάρτητη της πολυπλοκότητας της δομής του υδρολογικού μοντέλου.

Εφαρμογή μεθοδολογίας σε μεγάλη κλίμακα

Το επόμενο στάδιο της διερεύνησης αφορούσε την εφαρμογή της προτεινόμενης μεθοδολογίας σε ένα μεγάλο σετ λεκανών απορροής, σε μηνιαία κλίμακα προσομοίωσης. Συγκεκριμένα, η μελέτη αυτή πραγματοποιήθηκε επιλέγοντας 100 λεκάνες από τη βάση δεδομένων MOPEX (MOdel Parameter Estimation EXperiment), οι οποίες συνοδεύονταν από τα αντίστοιχα σετ δεδομένων των υδρολογικών διεργασιών. Η τοποθεσία των επιλεγμένων λεκανών απορροής σε όλη την έκταση των ΗΠΑ φαίνεται στο **Σχήμα 5**.

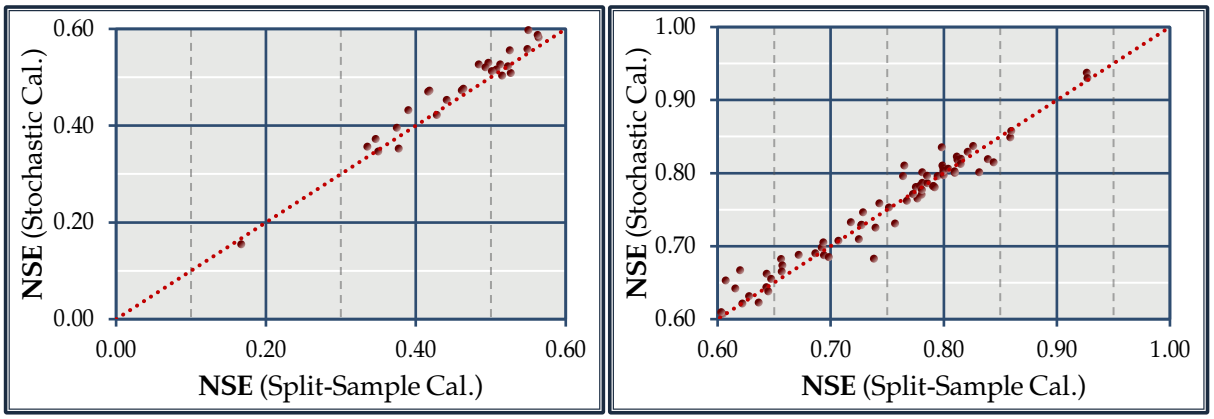
Όπως και στην προηγούμενη ανάλυση, έγινε χρήση ενός αριθμού υδρολογικών μοντέλων, προκείμενου να ενισχυθεί περαιτέρω η υπόθεση ότι η βαθμονόμηση με συνθετικά δεδομένα είναι ανεξάρτητη της πολυπλοκότητας του μοντέλου βροχής-απορροής που επιλέγεται. Ειδικότερα, τα μοντέλα που επιλέχθηκαν για τον σκοπό αυτό ήταν τα GR2M, Zygos4P και Zygos6P. Σημειώνεται ότι και σε αυτή την ανάλυση, για την περίπτωση ρύθμισης των μοντέλων με διαχωρισμό του ιστορικού δείγματος σε σετ βαθμονόμησης και επαλήθευσης, τα σετ αυτά είναι ίσα.



Σχήμα 5 | Γεωγραφική απεικόνιση των 100 επιλεγμένων λεκανών απορροής της βάσης δεδομένων MOPEX

Έχοντας ως γνώμονα την αποτελεσματική σύγκριση της επίδοσης (NSE) των υδρολογικών μοντέλων μεταξύ των εξεταζόμενων μεθόδων βαθμονόμησης, για την περίπτωση της στοχαστικής βαθμονόμησης επιλέγεται ως μέτρο σύγκρισης η επίδοση για το διάστημα επαλήθευσης του ιστορικού δείγματος, η οποία επιλέχθηκε κατά τη μεθοδολογία προσαρμογής του μοντέλου στα ιστορικά δεδομένα. Η απόφαση αυτή πηγάζει από το γεγονός ότι η περίοδος επαλήθευσης του δείγματος παρατηρήσεων περιέχει τα μόνα δεδομένα στα οποία, και στις δύο περιπτώσεις βαθμονόμησης, το εκάστοτε μοντέλο βροχής-απορροής δεν έχει προσαρμοστεί. Αντιθέτως, στην περίπτωση της ρύθμισης του μοντέλου με την κλασσική μέθοδο βαθμονόμησης, η επίδοση του μοντέλου κατά την περίοδο ρύθμισης είναι μεροληπτική, λόγω της υπερπροσαρμογής του μοντέλου στα δεδομένα αυτής της περιόδου.

Ενδεικτικά παρουσιάζονται στα ακόλουθα διαγράμματα (**Σχήμα 6**) τα αποτελέσματα της ανάλυσης με το μοντέλο Zygos4P. Συγκεκριμένα, η προσέγγιση που έχει προταθεί για βαθμονόμηση με συνθετικά δεδομένα, οδήγησε σε καλύτερη επίδοση για ένα σημαντικό ποσοστό των εξεταζόμενων λεκανών (67%), σε αντίθεση με τη μέθοδο βαθμονόμησης που το μοντέλο προσαρμόζεται στα ιστορικά δεδομένα.



Σχήμα 6 | Διαγράμματα διασποράς χαμηλής (αριστερά) και υψηλής (δεξιά) επίδοσης (NSE) για το μοντέλο Zygos4P, με εφαρμογή της μεθοδολογίας ρύθμισης με βάση τα ιστορικά δεδομένα (Split-Sample Cal.) και της στοχαστικής βαθμονόμησης (Stochastic Cal.)

Από τα αποτελέσματα της ανάλυσης μεγάλης κλίμακας συνάγεται το συμπέρασμα ότι για την πλειοψηφία των εξεταζόμενων υδροσυστημάτων, η μέθοδος στοχαστικής βαθμονόμησης αποδίδει καλύτερα από αυτή της βαθμονόμησης με τα ιστορικά δεδομένα, ανεξάρτητα από το υδρολογικό μοντέλο που επιλέχθηκε, καθιστώντας έτσι την βαθμονόμηση με συνθετικά δεδομένα ανεξάρτητη από την πολυπλοκότητα της δομής του υδρολογικού μοντέλου.

Όσον αφορά την απόκλιση μεταξύ των εκτιμημένων τιμών των παραμέτρων από κάθε μια από τις δύο μεθοδολογίες βαθμονόμησης, αυτή αποδείχθηκε ολοένα και πιο σημαντική, όσο η δομή του εξεταζόμενου μοντέλου γινόταν πιο πολύπλοκη.

Ολοκληρώνοντας, από την παραπάνω διερεύνηση το προτεινόμενο μεθοδολογικό πλαίσιο κρίνεται ως λειτουργικό. Παρ' όλα αυτά θα πρέπει να διερευνηθεί περαιτέρω και κάποιες πρώτες κατευθυντήριες γραμμές δίδονται αναλυτικά στο τέλος της παρούσας εργασίας.

Λέξεις – Κλειδιά: Υδροπληροφορική, Υδρολογικά Μοντέλα, Στοχαστικά, Στοχαστική Βαθμονόμηση

1

INTRODUCTION

1.1 Study Objectives

The aim of this study is to investigate the validity of a promising strategy regarding hydrological calibration, given synthetic forcing data, which are generated to preserve the essential statistics of the parent (i.e. historical) time series (probabilistic properties and dependence structure). To this end, the so-called *stochastic calibration* methodology is initially tested in the Loing river basin, as a proof-of-concept case study. During the research process, it was of significant importance to demonstrate that calibration with synthetic data is independent of the chosen model and time scale, thus (i) a number of different hydrological models were employed and (ii) across two different time scales, monthly and daily. Then, the study is extended to 100 catchments across USA, to further fortify the applicability of the proposed framework on a large scale.

1.2 Thesis Outline

To achieve these objectives, this study is structured as follows:

Chapter 2 contains a literature review in order to document the various outlines proposed for hydrological calibration, as well as the insight offered by several scientists, aiming toward a more improved calibration scheme.

Chapter 3 provides information about this new calibration approach and its advantages over conventional methods discussed in **Chapter 2**. The generation of the stochastic time series is a key part to the success of the method; hence, certain emphasis is placed on the requirements of the stochastic model.

Chapter 4 demonstrates the theoretical structure of the modelling tools used in this study, along with the calibration methods used, are also presented in this chapter.

In **Chapter 5**, the potential of the proposed calibration scheme is explored in the context of a real-world case study, by testing several lumped conceptual hydrological models of varying structure and across two time scales, monthly and daily.

Moreover, **Chapter 6**, entails a large-scale analysis, where the methodology is applied in 100 catchments of MOPEX database at a monthly time scale, by employing three different hydrological models.

Finally, **Chapter 7** summarizes the findings of the study conducted and concludes with recommendations for further research.

2

OVERVIEW OF CALIBRATION APPROACHES IN HYDROLOGICAL MODELLING

Hydrological models are calibrated by adjusting model parameters according to the characteristics of the area of interest. Afterwards, an evaluation process should follow to secure the predictive capacity of the identified model structure.

Traditionally, for this two-step procedure the modeller considers the temporal allocation of the available observed data in two sub-periods (i.e., calibration and validation period). This widely used practice among hydrologists was publicized by V. Klemeš [1986] and it is known as the split-sample test. Specifically, Klemeš formalized a four-level testing scheme for temporal and spatial data distribution: (i) the split-sample scheme, (ii) the proxy-catchment test, (iii) the differential split-sample test, and (iv) the proxy-catchment differential split-sample test.

As stated above, the split-sample approach is the most common scheme implemented for the calibration/validation procedure. This method is applied to catchments that adopt the concept of stationarity, which renders model parameters as time-invariant. If the available sample of data is of sufficient length to such a degree that allows for providing half of them for the calibration process, it should be divided into two equal segments, one for calibration and one for validation. Should the results from the two processes be in good agreement, the model structure is deemed acceptable for simulating the catchment's response.

Provided that the length of the available dataset is not adequate for a 50/50 splitting, this should occur in two different ways. For example, the first 70% of the available records should be allocated for calibration, while the remaining dataset for validation, then the last 70% of the whole sample for calibration and the first 30% for validation. In case the model performance for each validation period is similar and at the same time adequate, then the model meets the requirements to simulate measured runoff for the specific basin. In the opposite case, which seems to be rather common due to the length of data, the model should proceed to one of the following tests to qualify.

The proxy-catchment scheme should be applied for model calibration, in case the available data are insufficient for using the split-sample approach, or for ungauged catchments [Klemeš, 1986]. Specifically, if the runoff at the outlet of an ungauged river basin is to be

modelled, the modeller should select two gauged basins (e.g., A and B), in proximity with the one of interest. Model calibration should be implemented on basin A and validated against basin B, and vice versa. This will eventually result in two sets of adjusted model parameters; therefore, the modeller should first ensure that both sets lead to acceptable and similar model performance and then decide which one is best for use, depending on higher model performance, but also by comparing simulated data with any available historical data of the ungauged basin.

Another calibration approach introduced by Klemeš [1986] is the differential split-sample test, a method recommended whenever it is required to model streamflow in a gauged basin under different conditions from those prevailing over the observed flow data. Depending on the nature of the change for which the flow should be predicted, like a climate change scenario or change of environmental conditions, this test may differ. Specifically, Klemeš demonstrated the potential risk of using a model for simulating climate change effects without first undergoing a differential split-sample analysis.

Regarding the simulation of the impacts of a possible climate change, two periods with different values of the climate variables of interest, such as high and low average precipitation, should be selected. If the model is meant for simulating streamflow in a wet climate scenario, it should be adjusted using data from the part of the historical records, for which dry hydroclimatic conditions dominate, and assessed by data from the wet part of the available time series. This aims towards evaluating model transferability over time under various climate conditions. In case the model is intended to simulate flows for a dry climate scenario, the opposite process should be applied.

The fourth approach that Klemeš suggested was the proxy-catchment differential split-sample test, which is a combination of the latter two approaches described above. This method is considered as the most difficult of the four for the evaluation of a hydrological model, since it is used in case studies for which there are no available data for calibration and the model is intended to predict time-variant conditions, such as possible climate change scenarios. Moreover, this test should be applied for calibration in cases where the model is supposed to be both geographically and climatically (or land-use-wise) transposable.

In particular, if a model is intended for evaluation of the impact of climate change in an ungauged basin C, the test should have the following form: two gauged basins, A and B, with characteristics similar to those of basin C, are selected and temporal samples with different climatic variables of interest are identified in the observed data for each catchment, e.g., wet and dry hydrological regime. Then, for the dry climate case, the model is calibrated against the dry subset for the first test catchment and a validation test is run

using the contrasting set (e.g., the wet period data for the second catchment). Then, the model is calibrated on the wet sample for the second catchment and validated on the dry sample of the first test catchment. The model is considered adequate if the validation run errors for both catchments are acceptable and do not differ significantly.

Regarding the split-sample scheme, robustness of the optimized parameter set can be questioned when model performance for calibration period is significantly higher than that of the validation period. The selection of training data that are not representative of the dominant hydrological conditions, often results to this case, due to the over-fitting of the model structure to this period. As a consequence, the allocation data for model training and the length of those data is a pivotal decision.

A certain number of variants of the split-sample scheme have therefore proposed. Arnold *et al.* [2012] suggest the use of a long calibration dataset in order to include varying hydroclimatic conditions, whereas another strategy proposes that data should be allocated in such a manner as to encompass "unusual" events [Singh and Bárdossy, 2012]. What is more, the role of quality and quantity of the data to be used for model parameters adjustment has been first explored by Gupta and Sorooshian [1985], while presenting a suitable method for data selection. Attention has also been paid to the potential role of "soft" data and how these can contribute to the calibration process in conjunction with the "hard" data (e.g. observations) [Seibert and McDonnell, 2002].

Another variant for the split-sample approach is the odd/even method. Namely, the calibration procedure occurs on the odd years of the sample, while the validation dataset consists of the even years, or vice versa. This form of the split-sample test was developed in an effort to encapsulate any non-stationary trend information in the resulted parameter set. Notable applications of the method were accomplished by Arsenault *et al.* [2017], Essou *et al.* [2016] and Gowda *et al.* [2012], the last of whom secured that, by using this method on a dataset characterized by a wet first half and a dry last half, the model structure would be adjusted on the whole range of the available measured runoff values.

Deduced from the above discussion, the validation process is an integral part of the described operation. Thus, a data segment should always be allocated for this phase, resulting to important hydrologic information being entirely ignored. Consequently, hydrological researchers tried to face this challenge by proposing a general calibration scheme, either by testing all possible combinations of calibration-validation periods across sliding windows (e.g., Coron *et al.* [2012]) or by testing model performance in different sub-periods (e.g., Gharari *et al.* [2013]), for the purpose of providing a consistent performance in time. The studies of Singh and Bárdossy [2012] and Arsenault *et al.* [2018] went so far as to attempt an unconventional practice, to completely forego the validation

run and calibrate the model over the entire dataset. Another study regarding calibration against the full length of the available data was conducted by Razavi and Tolson [2013], who implemented a procedure to deal with demanding run time of computationally expensive hydrological models by calibrating against the full dataset, using surrogate-based methods on shorter periods.

Concerning the forecasting of the performance of a calibrated model, in which case the aforementioned forms of the differential split-sample test are deemed the appropriate for model calibration, a strong reliance of model parameters on the hydroclimatic conditions of the calibration period has been identified [Coron et al., 2011]. Moreover, there are well-grounded doubts whether a short sample of observations can provide an adequately calibrated model in order to predict sufficiently the impact of the altered climate conditions [Thirel *et al*., 2015b]. Several researchers (e.g., Brigode *et al*. [2013]; Coron *et al*., [2014]; Duethmann *et al*. [2020]) tried to improve model performance when projecting under different climate scenarios, by using longer data segments for calibration, but eventually with hardly any improvement on the robustness of these models.

The above mentioned drawbacks of the presented calibration tests show that there is a clear need for a new scientific approach which will combine calibration with longer datasets, while these data embed a wide range of plausible hydroclimatic conditions.

3

THE CONCEPT OF STOCHASTIC CALIBRATION

3.1 Methodology

For the classical split-sample strategy applied to hydrological calibration, a deterministic hydrological model is used to transform the hydrometeorological processes (precipitation, evapotranspiration, etc.) of a river basin to its response processes. The model is using a set of parameters to simulate the hydrological behavior of the catchment, and their values are identified through calibration. To this purpose, a specific time period is chosen from the observed records and the model parameters are adjusted according to performance criteria, for the modelled streamflow to better approximate the observed one.

Subsequently, the predictive capacity of the model is assessed on the remaining of the observed data (validation process), to ensure the model's adequacy for reproducing the hydroclimatic conditions of a period different than that of calibration.

The deterministic hydrological calibration-validation scheme is presented in **Figure 3.1**.

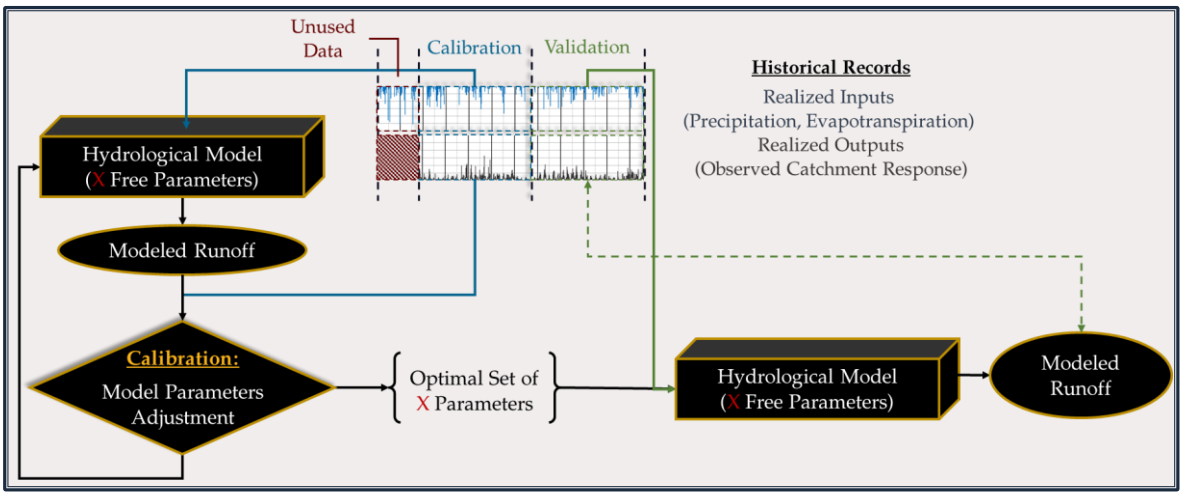


Figure 3.1 | Schematic representation of the classical split-sample scheme

Given that the available historical data may often not be representative of the catchment's hydrological regime and long-term hydroclimatic changes (a common pitfall in the case of data-scarce catchments), the typical split-sample structure may be proved to be deficient.

Moreover, discrepancies between the evaluation criteria for calibration and validation samples may indicate weak parameter identification. Thus, there is a need to extend the

temporal horizon chosen for calibration, so that the model can be trained over longer periods and capture the catchment dynamics to a greater extent. However, a part of the whole dataset must be allocated for validation purposes, hence precious hydrologic information encapsulated in the associated dataset is sacrificed.

As can be seen in the previous outline of this approach (**Figure 3.1**), only the overlapping periods of the observed input and output data can be exploited for model calibration and validation, ergo the remaining period of data is not accounted for. This is a commonly encountered scenario, since the length of rainfall records often exceeds that of the runoff observations.

In conclusion, the separation strategy, as well as the length of the calibration period, can pose a serious problem for hydrological modelling applied based on the split-sample rationale.

Therefore, a new calibration approach for parameter identification was devised [Efstratiadis *et al.*, 2021], which utilizes stochastic simulation, hence called *stochastic calibration*. The proposed framework for stochastic calibration is a simple conceptual approach on the basis of the conventional calibration-validation logic. Specifically, model calibration is accomplished by using long synthetic data as inputs, while the adjusted model structure is validated against the full historical records. The conceptual scheme of stochastic calibration is provided below (**Figure 3.2**).

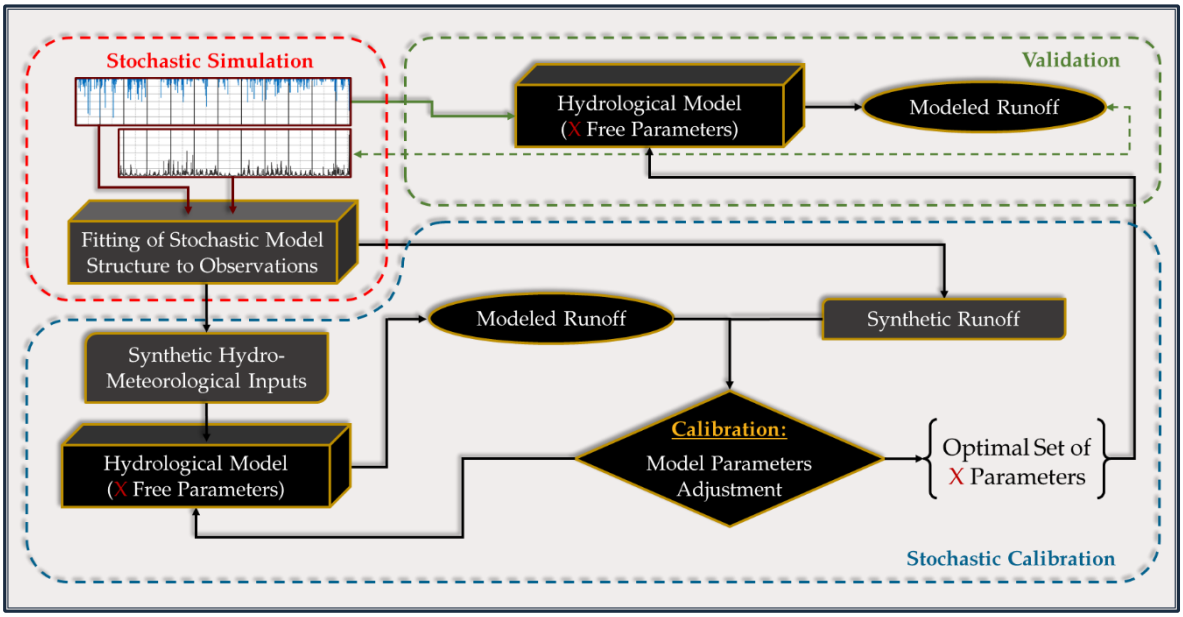


Figure 3.2 | Schematic representation of the conceptual structure of stochastic calibration

For the implementation of this strategy, an additional component is required, which precedes the calibration procedure. A stochastic model generating synthetic time series,

stochastically resembling the observed ones. Within those synthetic data should be reproduced all statistical information regarding the full hydroclimatic regime of the basin. Such information is integrated in the realized input and output data and is usually hard to identify within the, typically short, period of observations. What is more, this is further intensified since a significant part of this information is accounted for validation purposes within the split-sample calibration scheme.

Therefore, the drawbacks stemming from the application of the classical split-sample scheme are thereby eliminated, since the substantially longer synthetic sample, that is expected to describe the full hydroclimatic regime, distributed for model calibration will eventually lead to more robust parameters and stable predictive capacity. Furthermore, the predictive capacity of the model is now evaluated against the full set of observed data, hence the validation dataset is also extended. Given these conditions, the dilemma of which part of data to allot for calibration and which for validation, does not exist anymore.

3.2 Insights to the Stochastic Simulation Prerequisites

Throughout the hydrological simulation of hydrosystems, deterministic models are often proved incompetent to forecast hydrometeorological processes, thus all fluxes are handled as random variables and they are generated with the aid of stochastic models. It is a frequently encountered practice to replace the historical time series, which are usually of short length, with synthetic ones, being used as drivers for deterministic models employed in water resources planning and management applications. The conjunctive use of deterministic and stochastic models does not pose any restrictions regarding the temporal horizon of interest, since the historical records, which usually have an insignificant probability of realization in the future due to their short length, are substituted by longer data (e.g., with length of hundreds or thousands of years), statistically equivalent to the historical ones. Moreover, it is feasible to draw safe conclusions about the reliability of the system with satisfactory accuracy.

The principal requirement for generating statistically consistent synthetic inflows is the preservation of all the essential statistical characteristics of the observed data, at multiple time scales, in terms of the marginal statistics up to third order (i.e., mean, variance, skewness) and the joint second order properties (auto- and cross-correlations) [Matalas and Wallis, 1976].

However, for the stochastic calibration approach, this prerequisite is not sufficient, since these properties of the parent data describe partially only the marginal and dependence patterns of the hydrological processes to be reproduced, whereas, in case of some quite common circumstances, limited dependence patterns may appear in the synthetic data,

which are unrealistic and inconsistent with the historical records; a case known as *envelope behavior* [Tsoukalas *et al.*, 2018b].

Therefore, a consistent, and at the same time robust, stochastic calibration framework should satisfy the following specifications throughout the synthetic data generation procedure:

1. Concise representation and simulation of the hydrometeorological inputs and outputs, by reproducing all the probabilistic aspects of these time series, instead of a blind reproduction of the statistical characteristics of the observed data. This is achieved by assigning a suitable distribution model to each process, according to variable type and time scale of study [Tsoukalas *et al.*, 2019, 2020]. This requires stochastic simulation schemes which consider the main peculiarities of hydrometeorological processes, i.e., non-Gaussianity, intermittency, periodicity and time-irreversibility ([Efstratiadis *et al.*, 2014]; [Tsoukalas *et al.*, 2018a, 2019, 2020]; [Kossieris *et al.*, 2019]; [Koutsoyiannis, 2020]). What is more, should the purpose of hydrological simulation be the simulation of extreme events (e.g., floods), it is important to assess the tail behavior of the distribution models, that is, the upper part of a probability distribution, which governs both the magnitude and frequency of extreme events.
2. Representation of auto-dependencies (short- or long-range) and cross-dependencies, which constitute statistical indicators of the cause-effect relationships across the hydrological cycle. Specifically, short-range dependence entails an exponential autocorrelation structure which decreases after few time lags, a phenomenon traditionally referred to as "memory", whereas long-range dependence, also known as "long-memory" or long-term persistence, signifies an auto-dependence structure that strongly extends for a large number of time lags. The latter relates to particular patterns of similar extreme events (wet or dry) which tend to present a clustering behavior across all scales [Koutsoyiannis, 2013]. Regarding the cross-dependence, those determine the statistical interdependencies between the input and output hydrometeorological processes; hence a large cross-dependence value corresponds to a strong relation of the output process to the input one.
3. Multi-scale consistency, to ensure the reproduction of the probabilistic and stochastic behavior of the modelled processes not only at the scale of the hydrological simulation (e.g., daily, monthly), but also at higher (coarser) time scales (e.g., annual).

4

HYDROLOGICAL MODELLING TOOLS

In order to test the functionality of the proposed framework for stochastic calibration, a number of conceptual hydrological models with conceptual structure of varying complexity have been either developed or used. Each one of them is a bucket-type model with a lumped schematization, which uses a set of mathematical equations to simulate the main hydrological mechanisms at the catchment scale.

Both hydrological models Zygos4P and Zygos6P were configured in R-environment (the code developed for both models is presented in **Appendix A: R script for "Zygos4P" and "Zygos6P" hydrological models**), whereas the aforementioned GR models have been already implemented within the R–package *airGR* [Coron *et al.*, 2017; Coron *et al.*, 2021].

The conceptual structure of each of the five models used in this study, as well as the equations which regulate each one of them, are presented in the following sections **4.1** – **4.5**.

4.1 Hydrological Model Zygos4P

For the purposes of this study, a four-parameter lumped water balance model was devised, based on a simplified version of Zygos model [Kozanis & Efstratiadis, 2006], with conceptual structure shown in **Figure 4.1**. The input data are areal precipitation (P), in the form of rainfall, and potential areal evapotranspiration (PET). Moreover, the initial conditions of the model should be determined in order to run the water balance model, those being expressed through the level of the soil moisture reservoir (S_0) and the groundwater reservoir (G_0) at the beginning of the simulation. All different hydrological fluxes are expressed in units of water depth per unit time (i.e., mm/month or mm/day), while storages are expressed in terms of water depths (i.e., mm).

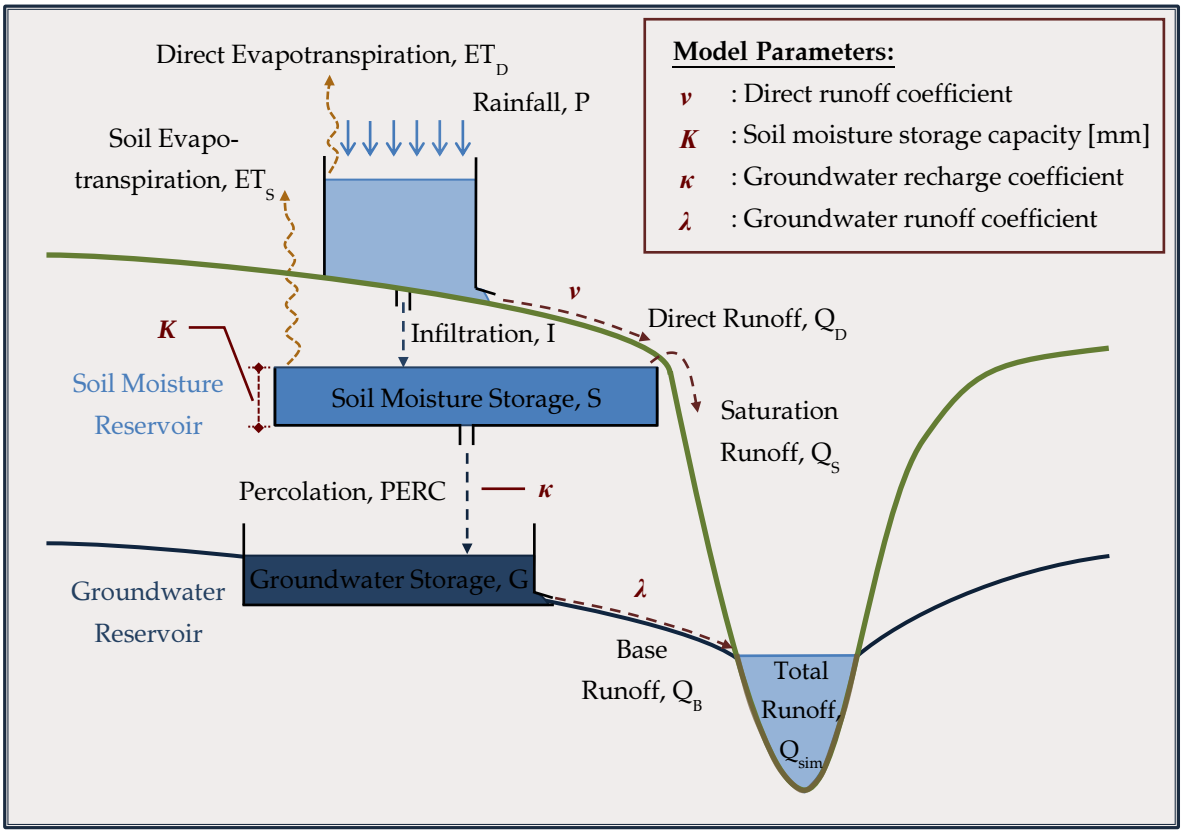


Figure 4.1 | Schematic representation of the conceptual structure of the Zygos4P

The proposed model structure defines four parameters:

- the impervious area percentage v, representing the rainfall part which flows directly on the soil surface, without infiltrating in the soil;
- the capacity K [mm] of the soil moisture reservoir, which corresponds to the maximum storage potential of the unsaturated zone;
- the groundwater recharge coefficient κ, reflecting the percentage of water which is channeled into the saturated zone from the unsaturated zone;
- the groundwater runoff coefficient λ, which is the release rate of groundwater to the hydrographic network as base flow.

The model scheme conceives direct runoff, $Q_D(t)$, as a percentage v of the rainfall, $P(t)$, while the remaining rainfall fulfils by priority the PET demand at the specific time step t, thus producing direct evapotranspiration, $ET_D(t)$, i.e.:

$$Q_D(t) = v\, P(t) \tag{4.1}$$

$$ET_D(t) = \min[\, PET(t) \,;\, P(t) - Q_D(t) \,] \tag{4.2}$$

The remainder enters to the soil moisture reservoir, which represents the unsaturated zone of the soil. Therefore, its current storage is increased to:

$$S(t) = S(t-1) + I(t) \quad (4.3)$$

where I(t) is the infiltrated through the soil quantity.

As regards the processes in the unsaturated zone, the upward vertical outflow of the reservoir corresponds to the evapotranspiration through the soil, $ET_S(t)$, whereas the downward vertical outlet results from the water percolating until it reaches the saturated zone of the soil, which is modelled by approximating the zone as a second linear reservoir (groundwater reservoir). The soil evapotranspiration is dependent on the filling rate of the soil reservoir and is estimated via the empirical law of Thornthwaite:

$$ET_S(t) = S(t)\left(1 - e^{-\frac{PETD(t)}{K}}\right) \quad (4.4)$$

where $PETD(t) = PET(t) - ET_D(t)$ denotes the potential evapotranspiration deficit to be covered by the available soil moisture.

Percolation, PERC(t), is controlled by percentage κ:

$$PERC(t) = \kappa\, S(t) \quad (4.5)$$

In case the available soil moisture storage surpasses the capacity of the soil reservoir, K, the excess amount of water is transformed into saturation runoff, $Q_S(t)$, by means of overflow, i.e.:

$$Q_S(t) = \max[\, S(t) - K\, ;\, 0\,] \quad (4.6)$$

As mentioned above, the percolation component feeds the groundwater reservoir, thus rising its storage to:

$$G(t) = G(t-1) + PERC(t) \quad (4.7)$$

It should be noted that this is an unconstrained reservoir in terms of capacity and features a single horizontal outlet. This outlet models the base flow, $Q_B(t)$, which is controlled by percentage λ:

$$Q_B(t) = \lambda\, G(t) \quad (4.8)$$

The total runoff, $Q_{sim}(t)$, is calculated by summing all three runoff components (i.e., direct, saturation and base runoff), while the actual evapotranspiration losses, $ET_{act}(t)$, are estimated as the sum of the direct and soil evapotranspiration:

$$Q_{sim}(t) = Q_D(t) + Q_S(t) + Q_B(t) \quad (4.9)$$

$$ET_{act}(t) = ET_D(t) + ET_S(t) \quad (4.10)$$

Finally, based on the water conservation law, the storage in each of the two reservoirs is updated at the end of each time step according to the following equations, as they serve as the initial conditions for the following time step:

$$S(t) = S(t-1) + I(t) - ET_S(t) - PERC(t) - Q_S(t) \quad (4.11)$$

$$G(t) = G(t-1) + PERC(t) - Q_B(t) \quad (4.12)$$

4.2 Hydrological Model Zygos6P

Adopting the structure of Zygos4P scheme, a six-parameter lumped model was also developed (see **Figure 4.2**) by implementing three modifications.

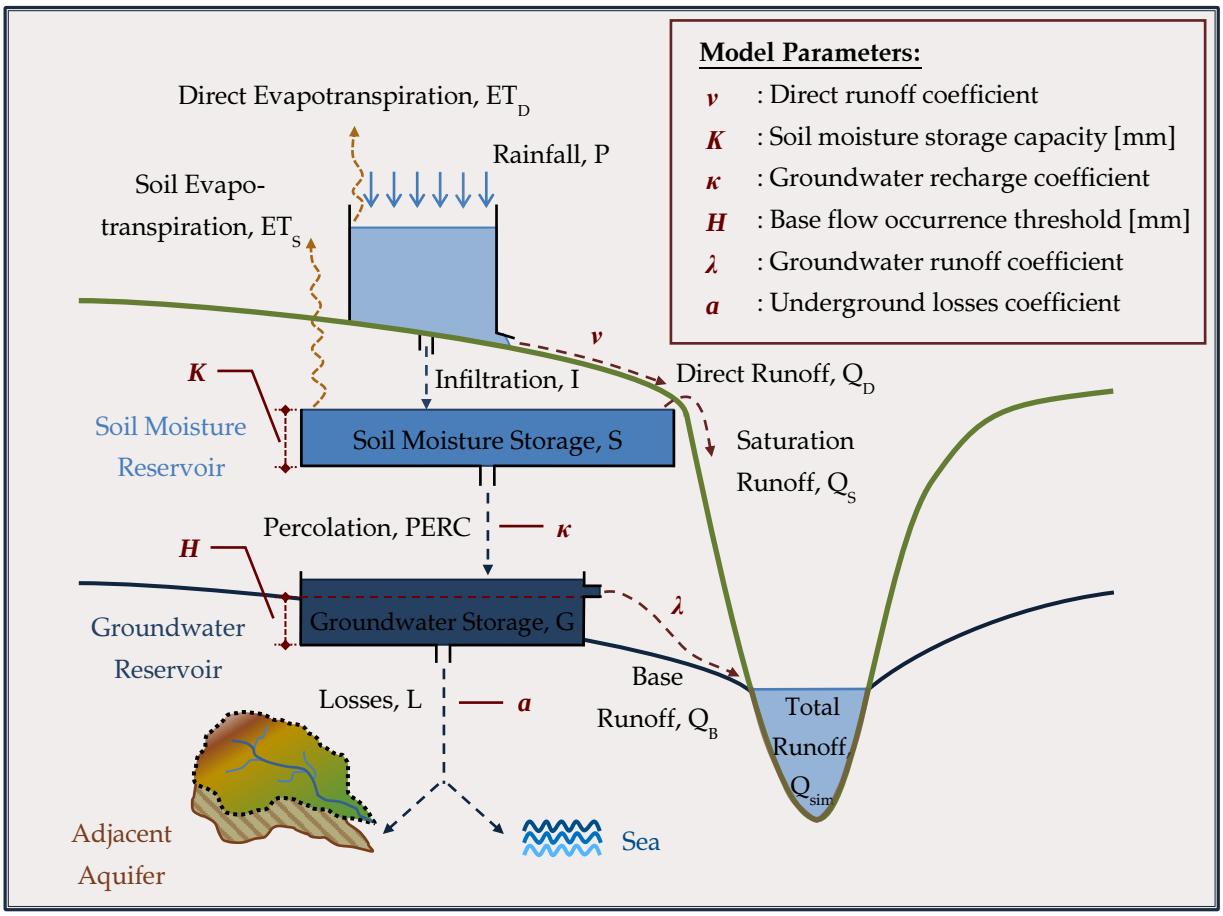


Figure 4.2 | Schematic representation of the conceptual structure of the Zygos6P

First, in order to enhance the model efficiency, a different approach was used to estimate direct evapotranspiration $ET_D(t)$ and direct runoff $Q_D(t)$. Specifically, the latter is considered a function of the soil moisture storage at the beginning of each time step, i.e., $S(t-1)$, and is calculated according to an empirical formula. Analytically:

$$ET_D(t) = \min\left[\frac{i^{[P(t)\neq 0]}}{N} P(t) ; PET(t)\right] \quad (4.13)$$

$$Q_D(t) = \left(P(t) - ET_D(t)\right)\left(\frac{S(t-1)}{K}\right)^{\nu} \quad (4.14)$$

where $i^{[P(t)\neq 0]}$ indicates the number of time steps t that a rainfall event occurred; N is the length of the rainfall data sample; and ν is a dimensionless model parameter, regulating

the direct runoff process.

Second, in this scheme, there is an elevation threshold, H [mm], for the groundwater reservoir, so that base runoff occurs. Thus, base runoff, $Q_B(t)$, is determined as a fraction, λ, of the groundwater storage above the threshold H, i.e.:

$$Q_B(t) = \max[\, \lambda \, (G(t) - H) \, ; \, 0 \,] \tag{4.15}$$

Third, the same reservoir is conceptualized with an additional (vertical) outflow, which corresponds to underground losses, L(t), to the sea or to the underground part of the adjacent river basins. These losses are calculated as a percentage a of the current groundwater storage:

$$L(t) = a \, G(t) \tag{4.16}$$

Except for the aforementioned components, all of the remaining hydrological processes are calculated in an identical way to that of the Zygos4P scheme.

Eventually, the above model contains six parameters, i.e.:

- the dimensionless coefficient v, regulating the direct runoff;
- the capacity K [mm] of the soil moisture reservoir, which corresponds to the maximum storage potential of the unsaturated zone;
- the groundwater recharge coefficient κ, reflecting the percentage of water which is channeled into the saturated zone from the unsaturated zone;
- the threshold H [mm] of the groundwater reservoir for base flow generation;
- the groundwater runoff coefficient λ, which is the release rate of groundwater to the hydrographic network as base flow;
- the fraction a of the groundwater storage that outflows to the sea or to adjacent aquifers.

4.3 Hydrological Model GR2M

The GR2M model belongs to the family of the GR models, a set of conceptual lumped hydrological models developed for specific time steps. The GR2M is a monthly time step model, which has been continuously developed to improve its efficiency. For this study, it was selected the most recent version of GR2M model [Mouelhi *et al.*, 2006]. The model inputs are catchment rainfall (P) and potential evapotranspiration (PET).

The GR2M model uses only two parameters:

- the maximum capacity X_1 [mm] of the production store;

- groundwater exchange coefficient X_2.

The conceptual scheme of this lumped model consists of two reservoirs, as illustrated in **Figure 4.3**. The first reservoir, denoted as S, represents soil moisture of the basin and controls the production function with a maximum capacity X_1, whereas the routing reservoir, denoted as R, controls the transfer function with a capacity of 60 mm.

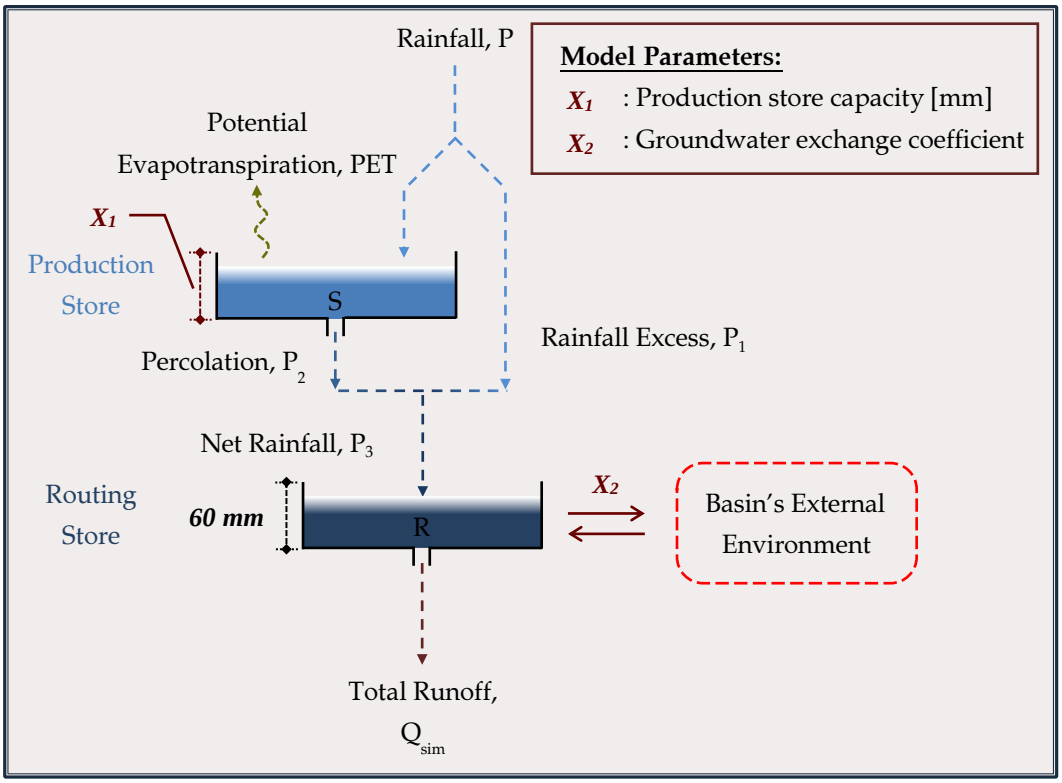


Figure 4.3 | Schematic representation of the conceptual structure of the GR2M model

At each time step, a fraction of the given rainfall, P, contributes to the soil moisture store. The new level of the production store, S_1, is estimated as follows:

$$S_1 = \frac{S + X_1\varphi}{1+\varphi\frac{S}{X_1}} \tag{4.17}$$

where $\varphi = \tanh(P / X_1)$.

In case the soil reaches its saturation point (X_1), rainfall excess, P_1, moves to the routing store and is obtained as:

$$P_1 = P + S - S_1 \tag{4.18}$$

Moreover, because of evapotranspiration losses, which are considered equal to PET, the soil reservoir reaches a new level, S_2, which is calculated as:

$$S_2 = \frac{S_1(1-\psi)}{1+\psi\left(1-\frac{S_1}{X_1}\right)} \tag{4.19}$$

where $\psi = \tanh(PET / X_1)$.

Additionally, the production store releases a quantity P_2, through percolation, attaining the following reservoir level at the end of this time step:

$$S = \frac{S_2}{\left[1+\left(\frac{S_2}{X_1}\right)^3\right]^{1/3}} \tag{4.20}$$

$$P_2 = S_2 - S \tag{4.21}$$

The percolated amount, P_2, is then combined with the rainfall excess, P_1, forming the net fraction, P_3, of the monthly rainfall, which is transferred to the routing store and added to the prior water content R. Analytically:

$$P_3 = P_1 + P_2 \tag{4.22}$$

$$R_1 = R + P_3 \tag{4.23}$$

Subsequently, a groundwater quantity is acquired or lost by the routing store due to water exchanges between the underground part of the catchment and its external environment. Specifically, in case the value of model parameter X_2 is positive, the routing store is supplied with water by the surrounding underground environment of the catchment, while in the opposite case there is a loss. Thus, the new routing store level is defined as:

$$R_2 = X_2 R_1 \tag{4.24}$$

Finally, the routing store releases the total runoff Q_{sim} at the catchment's outlet, which is calculated as:

$$Q_{sim} = \frac{R_2^2}{R_2 + 60} \tag{4.25}$$

The routing store level is updated at the end of the specific time step to:

$$R = R_2 - Q_{sim} \tag{4.26}$$

4.4 Hydrological Model GR4J

The GR4J [Perrin *et al.*, 2003], another model of the GR family, is a four-parameter daily lumped hydrological model. The model requires daily areal rainfall (P) and potential evapotranspiration (PET) as inputs. The schematic representation of GR4J model structure is given in **Figure 4.4**.

The GR4J model is controlled by the following four parameters:

- the maximum capacity X_1 [mm] of the production store;

- intercatchment groundwater flux X_2 [mm/day];
- the maximum capacity X_3 [mm] of the routing store;
- time base X_4 [days] for Unit Hydrographs.

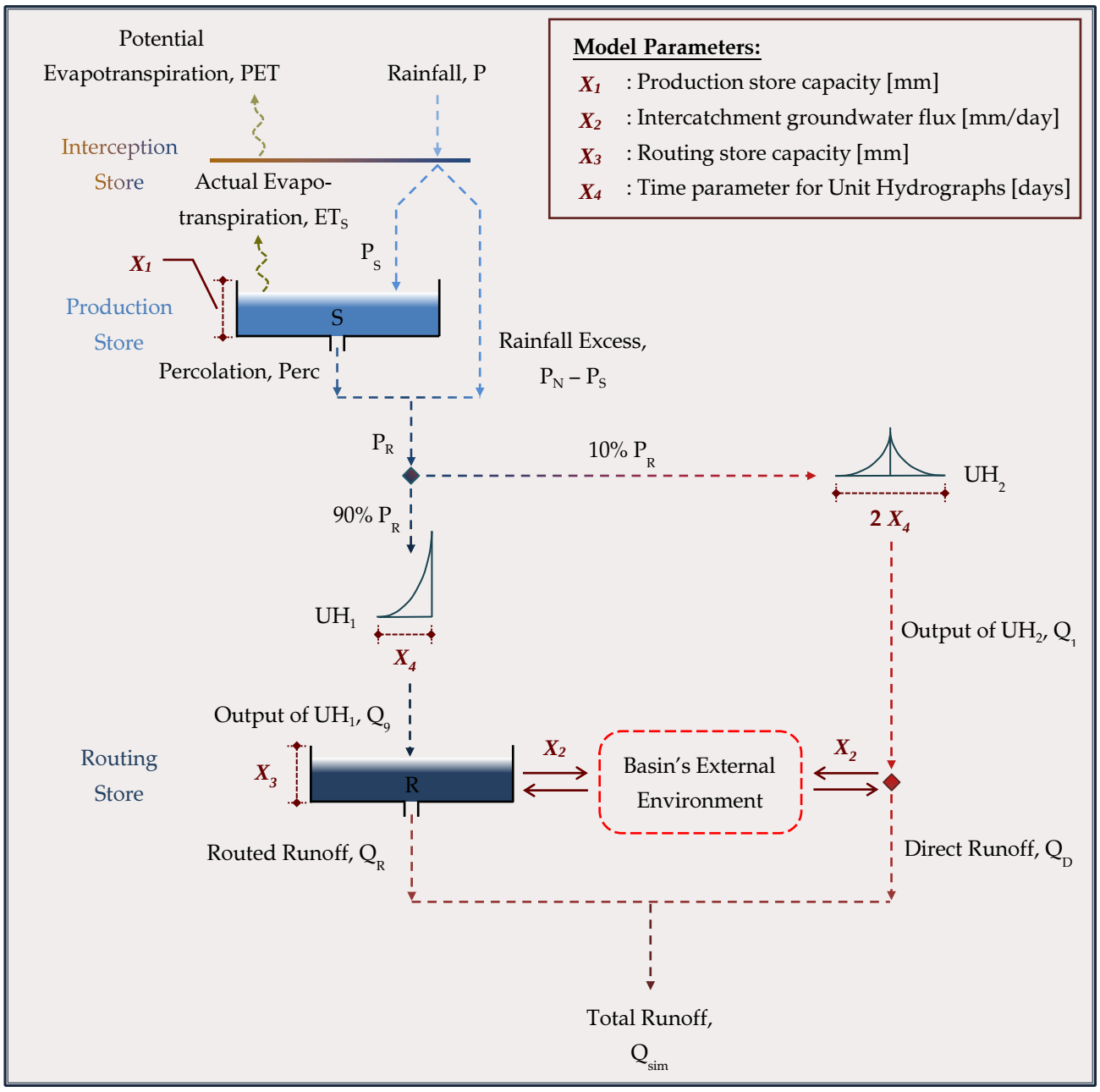


Figure 4.4 | Schematic representation of the conceptual structure of the GR4J model

As in the aforementioned model, the basin is vertically subdivided into two reservoirs (production and routing). The computational process begins with the deduction of the net rainfall (P_N) and net evapotranspiration capacity (ET_N) for the specific day. For this purpose, an interception storage of zero capacity is used where the model transforms the daily inputs into these quantities as follows:

$$P_N = \max[\, P - PET \,;\, 0 \,] \tag{4.27}$$

$$ET_N = \max[\, PET - P \,;\, 0 \,] \tag{4.28}$$

In case P_N is different from zero, a fraction P_S of this amount supplies the production store, which is calculated as a function of the store current level, S, its maximum capacity, X_1, and the amount of net rainfall, P_N:

$$P_S = \frac{X_1 \left(1 - \left(\frac{S}{X_1}\right)^2\right) \varphi}{1 + \frac{S}{X_1} \varphi} \tag{4.29}$$

where $\varphi = \tanh(P_N / X_1)$.

The remaining quantity of net rainfall, $P_N - P_S$, is directed towards the routing store.

In the other case, when ET_N is different than zero, the value of the actual evapotranspiration ET_S subtracted from the production store, is calculated as a function of the store current level, S, its maximum capacity, X_1, and the net evapotranspiration capacity, ET_N:

$$ET_S = \frac{S \left(2 - \frac{S}{X_1}\right) \psi}{1 + \left(1 - \frac{S}{X_1}\right) \psi} \tag{4.30}$$

where $\psi = \tanh(ET_N / X_1)$.

Then, the water level of the production store is updated to:

$$S_1 = S + P_S - ET_S \tag{4.31}$$

and a part of this quantity leaks from the reservoir as percolation, Perc, computed as:

$$Perc = S \left\{ 1 - \left[1 + \left(\frac{4}{9} \frac{S_1}{X_1} \right)^4 \right]^{-1/4} \right\} \tag{4.32}$$

Thus, the reservoir's content is eventually updated to:

$$S = S_1 - Perc \tag{4.33}$$

The water amount P_R to reach the routing store is given by:

$$P_R = Perc + (P_N - P_S) \tag{4.34}$$

This amount (P_R) is then split into two flow components:

- 90% of it, which is routed by a unit hydrograph UH_1 with time base equal to the value of model parameter X_4;

- 10% of it, which is routed by a unit hydrograph UH_2 with a time base equal to 2 X_4.

The ordinates $UH_1(i)$ and $UH_2(i)$ for each of those unit hydrographs are obtained from the corresponding S-curves (cumulative proportion of the input with time), whose ordinates are denoted as SH_1 and SH_2, respectively.

For each time step t, the outputs Q_9 and Q_1 of the two unit hydrographs UH_1 and UH_2, respectively, are calculated as follows:

$$Q_9(t) = 0.9 \sum_{k=1}^{l} [\, UH_1(k) \cdot P_R(t - k + 1) \,] \tag{4.35}$$

$$Q_1(t) = 0.1 \sum_{k=1}^{m} [\, UH_2(k) \cdot P_R(t - k + 1) \,] \tag{4.36}$$

where l = int[X_4] + 1 and m = int[2 X_4] + 2.

Similarly to the case of the GR2M model, a model parameter X_2 is used in order to simulate the groundwater exchanges between the catchment and its external environment. As in GR2M, a positive value of X_2 corresponds to water imports, whereas a negative value to water exports. Therefore, a groundwater exchange term F is calculated as:

$$F = X_2 \left(\frac{R}{X_3}\right)^{7/2} \tag{4.37}$$

where R is the routing store level and X_3 is its maximum capacity.

Then, the content of the routing reservoir is updated to:

$$R_1 = \max[\, R + Q_9 + F \,;\, 0 \,] \tag{4.38}$$

The routed runoff Q_R is estimated as:

$$Q_R = R_1 \left\{ 1 - \left[1 + \left(\frac{R_1}{X_3}\right)^4 \right]^{-1/4} \right\} \tag{4.39}$$

and the routing store level at the end of the time step is:

$$R = R_1 - Q_R \tag{4.40}$$

Like in the case of the routing reservoir, the flow component Q_1 interacts with the basin's outside environment, losing or gaining the same water quantity, F, hence the direct runoff component is obtained as:

$$Q_D = \max[\, Q_1 + F \,;\, 0 \,] \tag{4.41}$$

Finally, the delayed outflow Q_R combined with the direct runoff Q_D form the total runoff Q_{sim} at the catchment's outlet, i.e.:

$$Q_{sim} = Q_R + Q_D \tag{4.42}$$

4.5 Hydrological Model GR6J

The GR6J [Pushpalatha *et al.*, 2011], is one more lumped hydrological model of the GR family, driven with daily areal precipitation (P) and potential evapotranspiration time series (PET) and controlled by six parameters. The conceptual structure of the GR6J model is illustrated in **Figure 4.5**.

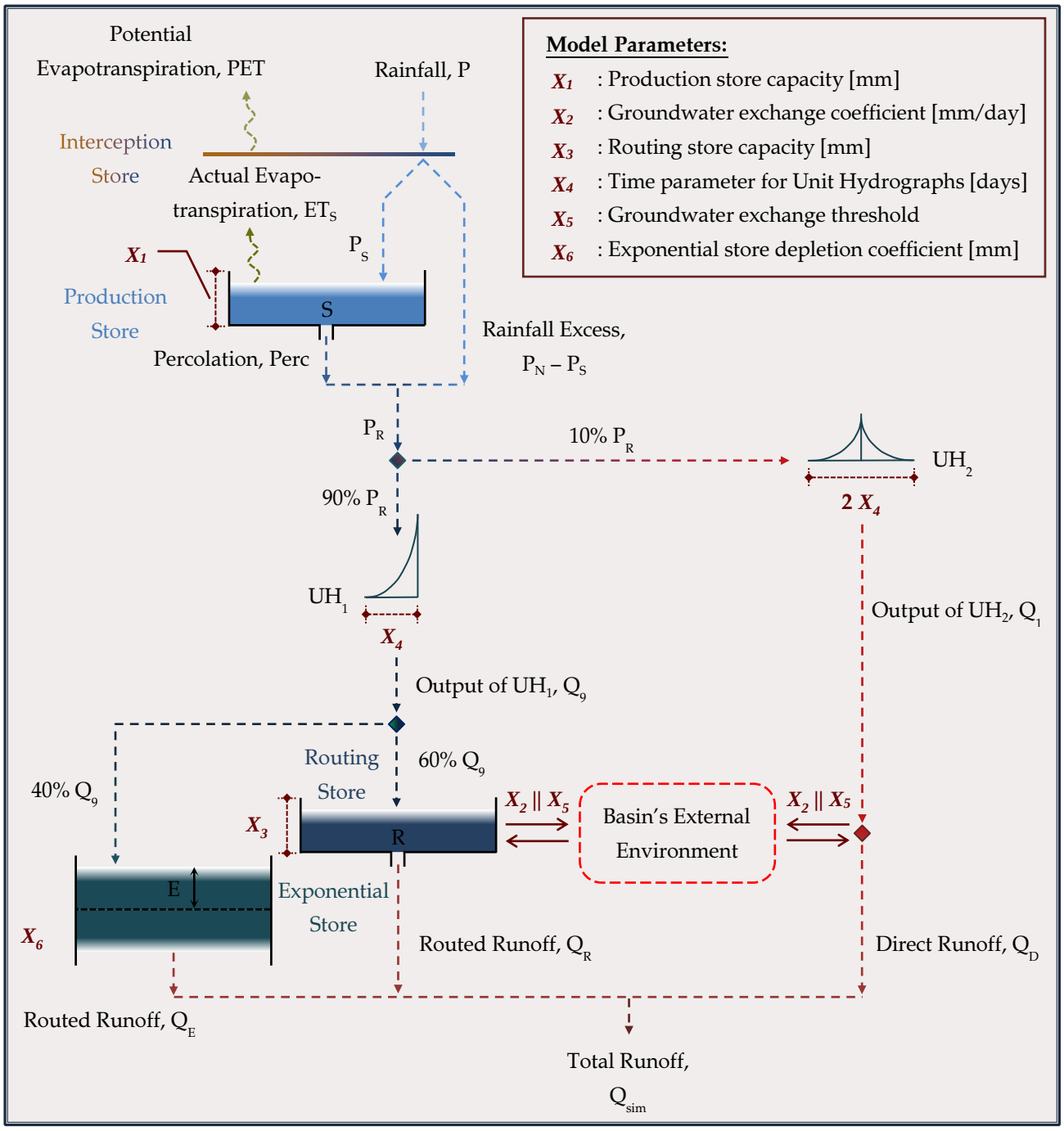


Figure 4.5 | Schematic representation of the conceptual structure of the GR6J model

The GR6J model is a modified version of the GR4J model, which was proposed for the improvement of modelling efficiency and specifically to achieve better low-flow representation. The modifications were progressively applied to the GR4J model. Le Moine [2008] has suggested a new groundwater exchange function, thus adding an additional parameter to the GR4J model, while Pushpalatha *et al*. [2011] introduced an exponential routing store, parallel to the existing one of the GR4J version, and controlled by a new parameter.

In particular, the GR6J model structure until the estimation of the two flow components Q_1 and Q_9 resulted from the two unit hydrographs (see **Figure 4.5**), remains the same as in the case of the GR4J model, hence the mathematical equations developed by Perrin *et al*. [2003] to describe the hydrological processes until that point (i.e., Eq. (4.27) to Eq. (4.36)) continue being valid for the GR6J model.

The water exchanges function F was modified to encapsulate possible changes in the direction of the groundwater exchange within the simulated year, based on the comparison of the current routing reservoir level (R_1) with the value of a dimensionless threshold parameter X_5, i.e.:

$$F = X_2 \left(\frac{R}{X_3} - X_5 \right) \tag{4.43}$$

where X_2 and X_3 are the respective model parameters presented in the section **4.4**.

In addition, the flow component Q_9 is divided into two parts. Sixty percent of it is directed towards the routing store, whereas the remaining 40% of the component is routed through the exponential store. The content of the routing and the exponential store are updated to:

$$R_1 = \max[\, R + 0.6\, Q_9 + F \,;\, 0 \,] \tag{4.44}$$

$$E_1 = E + 0.4\, Q_9 \tag{4.45}$$

The routed runoff Q_R is calculated by Eq. (4.39) and the routing reservoir level R at the end of the daily time step is updated according to Eq. (4.40).

As mentioned above, the exponential store is controlled by a model parameter, X_6, which acts as a base level in the reservoir. The routed runoff Q_E is estimated as a function of the current store level E_1 and this parameter.

Finally, the direct runoff Q_D is computed as stated in Eq. (4.41) and the total runoff Q_{sim} is obtained as follows:

$$Q_{sim} = Q_E + Q_R + Q_D \tag{4.46}$$

Therefore, the six parameters of the GR6J model are:

- the maximum capacity X_1 [mm] of the production store;
- intercatchment groundwater flux X_2 [mm/day];
- the maximum capacity X_3 [mm] of the routing store;
- time base X_4 [days] for Unit Hydrographs;
- groundwater exchange threshold X_5;
- the exponential store depletion coefficient X_6 [mm].

4.6 Calibration Algorithms and Criteria

Not all of the above hydrological models share a common calibration algorithm. Specifically:

- The parameter sets for both Zygos4P and Zygos6P models are estimated for all the study catchments using the Evolutionary Annealing-Simplex (EAS) optimization method [Efstratiadis & Koutsoyiannis, 2002], a hybrid scheme that merges the strengths of both local and global search. EAS optimization algorithm has already been implemented in R by P. Kossieris.
- The calibration of all three of the GR models was employed with the technique proposed by Michel [1991], an algorithm that also combines a local and a global approach and is already implemented in *airGR* R–package [Coron *et al.*, 2017; Coron *et al.*, 2021].

To evaluate the model performance, the measure used was Nash-Sutcliffe efficiency coefficient (NSE) [Nash & Sutcliffe, 1970]:

$$\mathrm{NSE} = 1 - \frac{\Sigma_{t=1}^{N}\left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\Sigma_{t=1}^{N}\left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2} \tag{4.47}$$

where $Q_{obs}(t)$ and $Q_{sim}(t)$ are the observed and simulated discharges, N is the length of the sample, and the overbar indicate mean value of the runoff records for this period.

Table 4.1 summarizes the parameters of each hydrological model, along with their ranges of variation, as obtained from the literature or specific applications of the models.

Table 4.1 | Parameters of conceptual rainfall-runoff models and their corresponding ranges of variation

Model	Parameter	Unit	Range	Reference
Zygos4P	v	-	[0 , 1]	Ranges deduced after numerous simulations
	K	mm	[0 , 1000]	
	κ	-	[0 , 1]	
	λ	-	[0 , 1]	
Zygos6P	v	-	[0.1 , 2.5]	
	H	mm	[0 , 300]	
	a	-	[0 , 1]	
GR2M	X_1	mm	[0 , 1500]	Mouelhi *et al.*, 2006
	X_2	– **or** mm/day	[-10 , 10]	
GR4J	X_3	mm	[1 , 500]	Perrin *et al.*, 2003
	X_4	days	[0.5 , 8]	
GR6J	X_5	-	[-4 , 4]	Le Moine, 2008
	X_6	mm	[0.5 , 20]	Pushpalatha *et al.*, 2011

5

PROOF OF CONCEPT

To assess the validity of the stochastic calibration scheme, it was initially implemented on a specific study basin, the Loing river basin, by examining two cases: one that the scheme is employed at a monthly time scale, and another at a daily time scale. For this comparative analysis between the split-sample scheme and the stochastic calibration, a code was developed in R-environment. For brevity purposes, only the code regarding the "Zygos6P" model is presented (**Appendix B: R script for Monthly Time Scale Analysis with "Zygos6P" model**).

5.1 Study Area and Data

The Loing catchment covers an area of 3900 km^2 and has a mean elevation of 148 m [Rebolho *et al.*, 2018]. The geographical representation of the basin is displayed in **Figure 5.1**. The boundary and hydrographic network of the catchment were taken from European Catchments and Rivers Network System (ECRINS). The mapping of the catchment was conducted using the ESRI software ArcGIS.

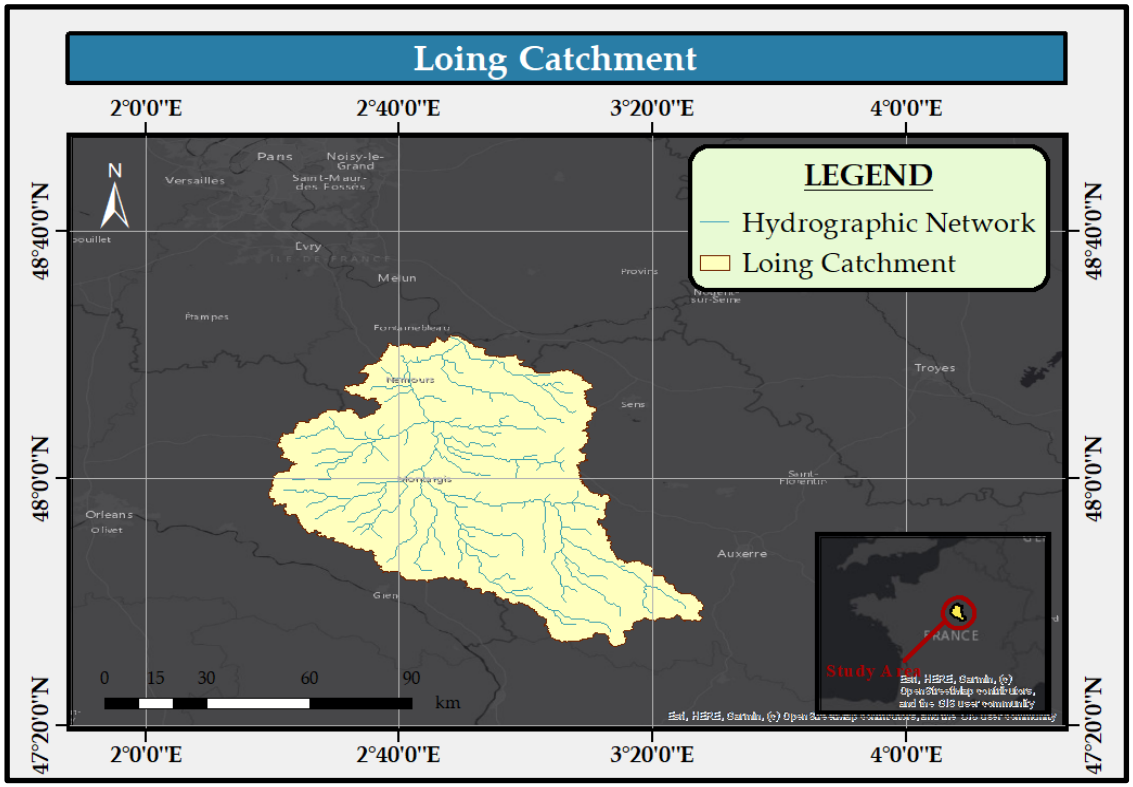


Figure 5.1 | Geographical representation of the Loing catchment

The available hydrological data (precipitation, potential evapotranspiration, runoff) are of daily time scale and extend over a period of 58 years (from 01/08/1958 to 31/07/2016) (**Figure 5.2**). The data were obtained from a previous study conducted by Rebolho *et al.* [2018] in the same catchment.

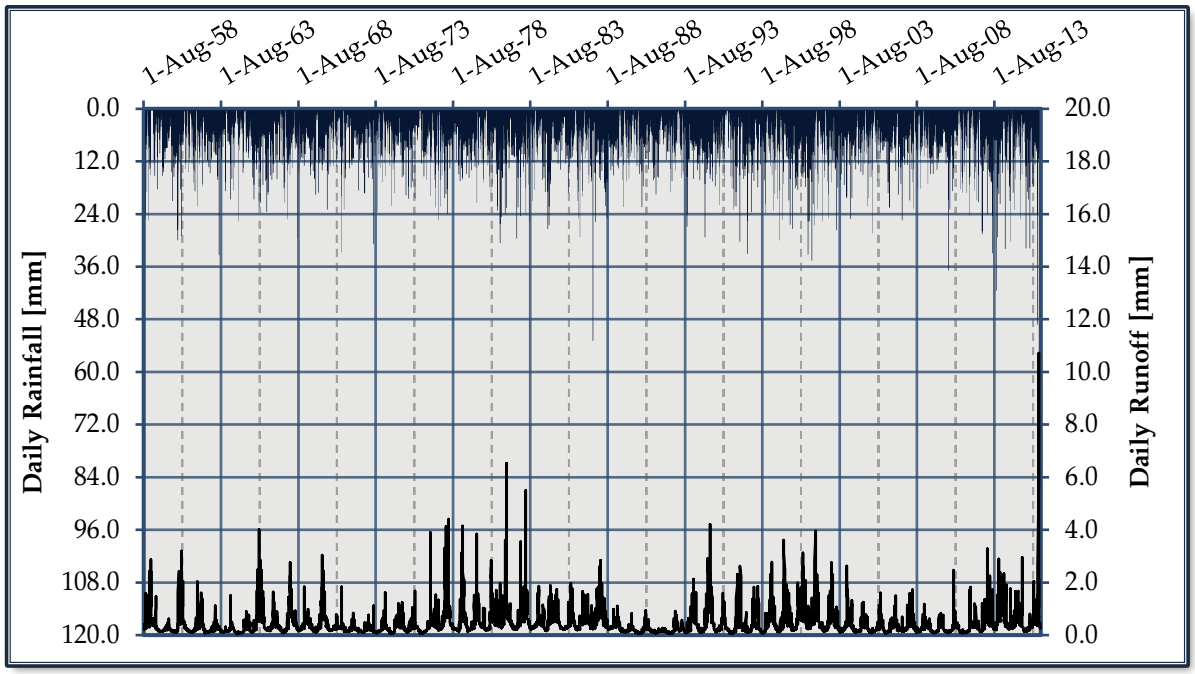


Figure 5.2 | Historical daily rainfall and runoff time series (August 1st, 1958 to July 31st, 2016)

5.2 Case Study A: Monthly Time Scale Analysis

For this specific study case, the hydrological models Zygos6P and GR2M were chosen to model monthly runoff at the outlet of the basin. Thus, the daily observed data are aggregated at the monthly scale and the resulting time series is depicted in **Figure 5.3**.

For each model calibration, the classical split-sample scheme is initially employed, by dividing the historical records into two equal subsets of length 29 years each, for calibration and validation.

Regarding the synthetic time series, these were generated via the *anySim* R-package [Tsoukalas et al., 2020], specifically designed for the simulation of non-Gaussian behavior, which characterizes hydrometeorological processes, apart from other significant peculiarities (periodicity, intermittency, and auto- and cross-dependence). This stochastic time series generator abides by the requirements regarding stochastic simulation (Section **3.2**) and it will be demonstrated below for this case study. For the sake of brevity, the respective process will be omitted for the case of generation of daily synthetic data.

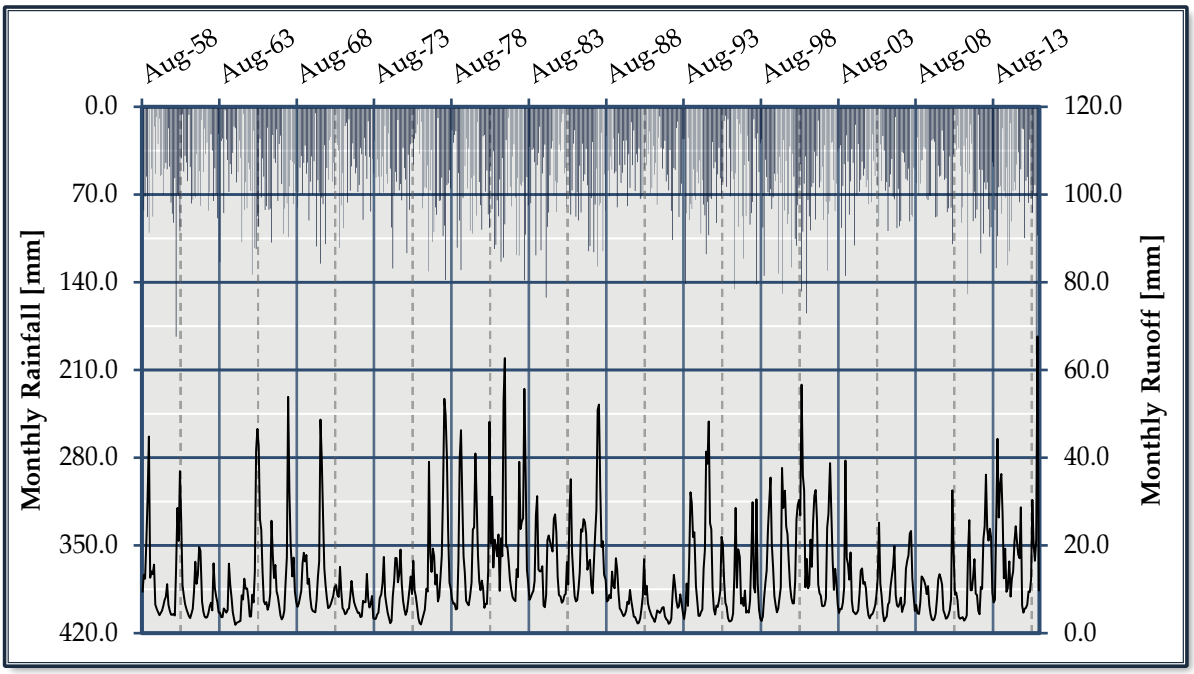


Figure 5.3 | Historical monthly rainfall and runoff time series (August 1958 to July 2016)

Table 5.1 displays the key statistics of historical rainfall and runoff data (mean, standard deviation, skewness, auto- and cross-correlations), to be reproduced within the synthetic time series that are used in stochastic calibration.

The stochastically simulated time series extend over a time horizon of 1000 years are presented in **Figure 5.4**.

Table 5.1 | Key statistical information of observed rainfall and runoff data and their lag-0 cross-correlation coefficient at the monthly and annual time scales

Rainfall [mm]	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Annual**
Mean	67.5	66.6	69.1	61.8	53.0	55.3	53.5	68.3	57.4	56.7	57.0	56.4	722.5
St. deviation	36.9	27.0	30.2	29.6	28.2	30.5	31.2	31.5	26.2	30.7	29.9	35.6	132.3
Skewness	0.603	0.554	0.488	0.508	0.537	0.856	0.558	1.019	0.458	1.111	0.217	0.865	0.131
Lag-1 correl.	0.162	-0.016	-0.173	0.035	-0.058	0.197	0.102	-0.031	0.296	0.135	0.108	-0.231	0.126
Runoff [mm]													
Mean	7.9	10.7	17.0	22.2	21.3	20.1	16.1	13.6	9.6	6.6	5.8	5.4	156.4
St. deviation	5.0	7.3	10.5	13.8	12.7	11.3	9.9	7.2	8.8	3.6	3.6	2.3	66.2
Skewness	1.769	2.310	1.177	0.962	0.820	1.345	1.659	0.95	5.284	1.402	2.404	1.044	0.357
Lag-1 correl.	0.713	0.786	0.576	0.672	0.514	0.717	0.683	0.746	0.438	0.455	0.833	0.703	0.483
Lag-0 correl.	0.616	0.392	0.610	0.681	0.748	0.758	0.625	0.436	0.394	0.532	0.330	0.202	0.841

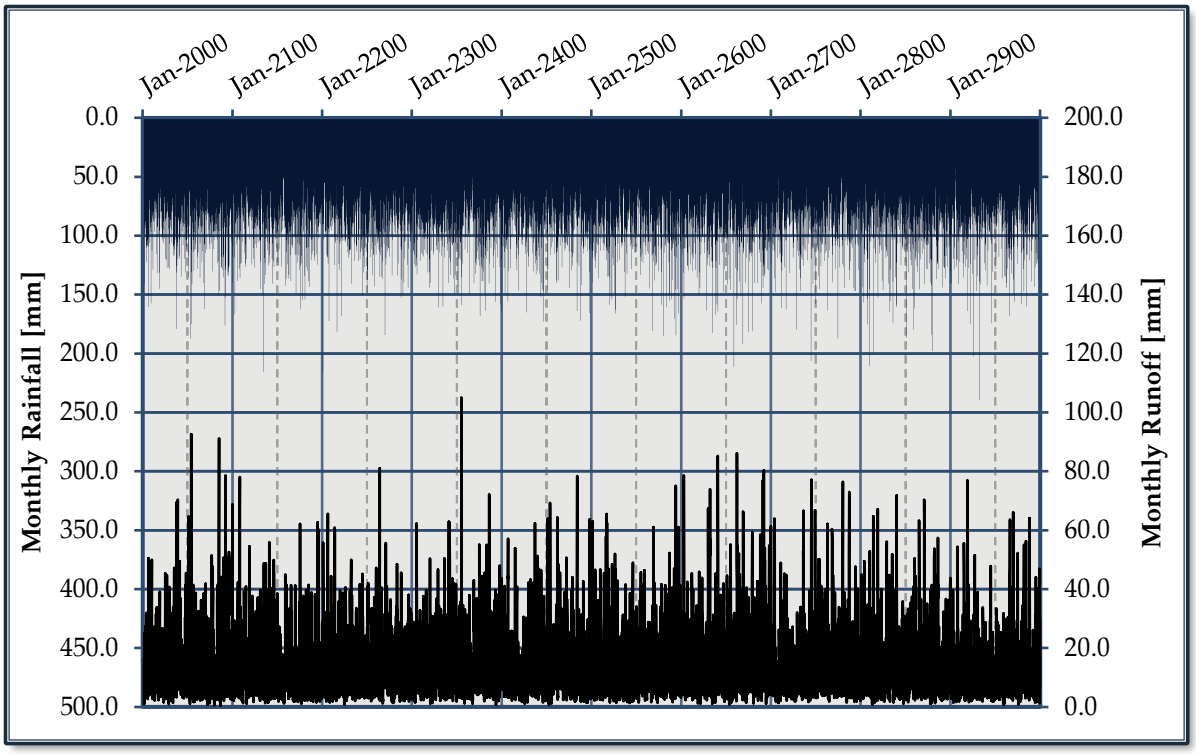


Figure 5.4 | Synthetically generated rainfall and runoff time series (randomly selected window of 1000 years)

Figure 5.5 summarizes the ability of the stochastic model to preserve the lag-0 cross-correlation among the two simulated processes (rainfall and runoff).

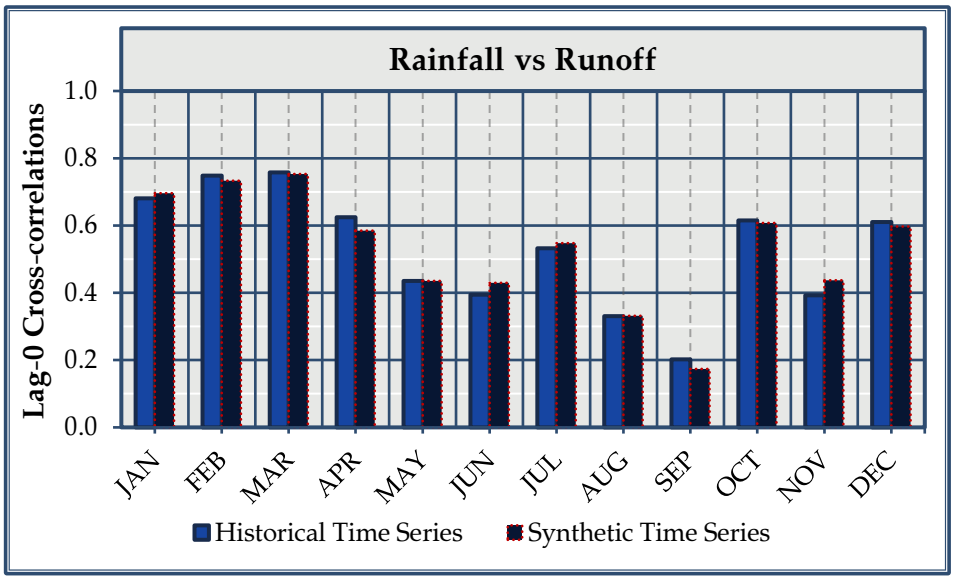


Figure 5.5 | Comparison of the monthly historical and simulated lag-0 cross-correlations

Furthermore, **Figure 5.6** illustrates that the synthesis scheme ensured the reproduction of dependency patterns that are much extended than the observed ones. These are expected to represent the full hydroclimatic regime of the basin, which cannot be traced in the case

of the observed data, due to their limited length. Actually, while using the split-sample calibration approach, only half of this information is accounted for.

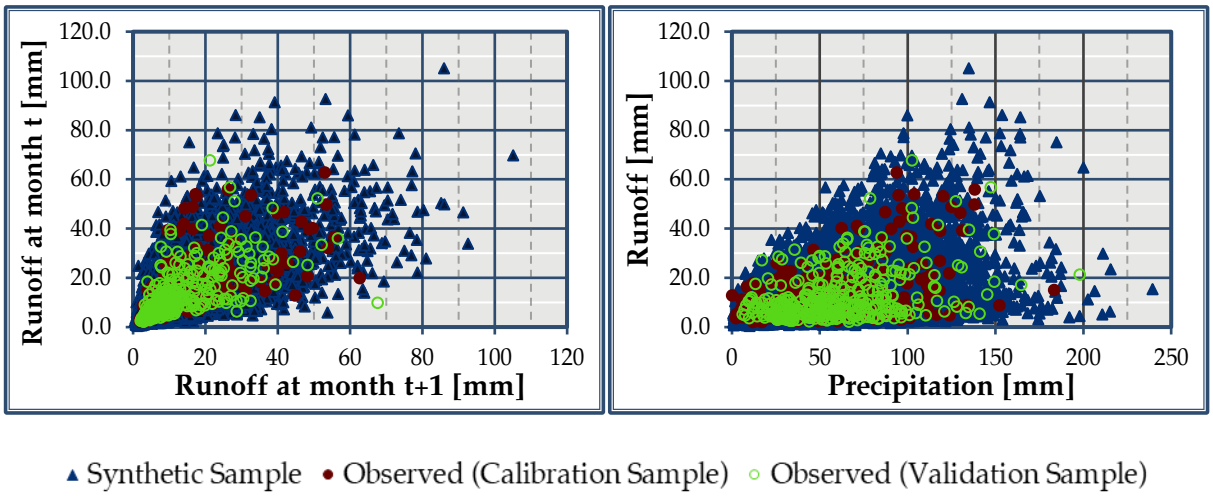


Figure 5.6 | Auto-dependency patterns among runoff data between subsequent months (left) and cross-dependency patterns between rainfall and runoff (right), derived from the observed data (split into two periods) and the synthetic ones (12 000 values)

It should be noted that for the estimation of the evapotranspiration input data were employed the mean monthly values of the historical sample.

Consequently, according to the stochastic calibration framework, the chosen models are fitted against the synthetic time series of length 1000 years, and their structure is validated against the full sample of observations (58 years).

Table 5.2 summarizes the results of the monthly time scale analysis using Zygos6P model.

Table 5.2 | Results summary for Zygos6P hydrological model calibration, employing the split-sample approach and stochastic calibration at a monthly scale

	Model performance evaluation			**Optimized Zygos6P parameter values**					
Split-Sample Approach	**NSE_{Cal}**	**NSE_{Val}**	**NSE_{Tot}**	***v***	***K* [mm]**	***κ***	***H* [mm]**	***α***	***λ***
	0.837	0.836	0.836	1.763	235.0	0.184	60.8	0.120	0.142
Stochastic Calibration	**NSE**			***v***	***K* [mm]**	***κ***	***H* [mm]**	***α***	***λ***
	0.836			1.835	234.2	0.180	44.2	0.157	0.173
Abbreviations	**NSE_{Cal}**: Nash-Sutcliffe efficiency metric in **calibration** period **NSE_{Val}**: Nash-Sutcliffe efficiency metric in **validation** period **NSE_{Tot}**: overall Nash-Sutcliffe efficiency metric **NSE**: overall Nash-Sutcliffe efficiency metric								

It is evident that the results of calibration using the split-sample approach are in good agreement with those of the stochastic calibration scheme; a fact also confirmed by visual inspection of **Figure 5.7** and **Figure 5.8**, where in both cases modelled runoff approximates well the runoff observations.

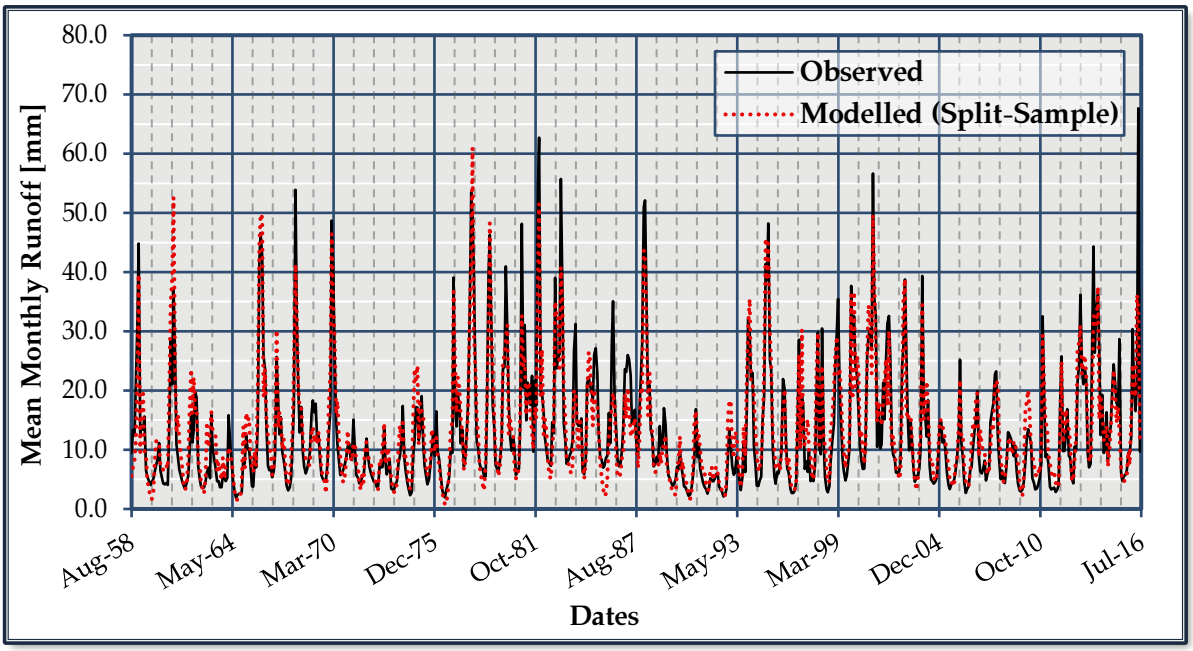


Figure 5.7 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach – **Zygos6P**)

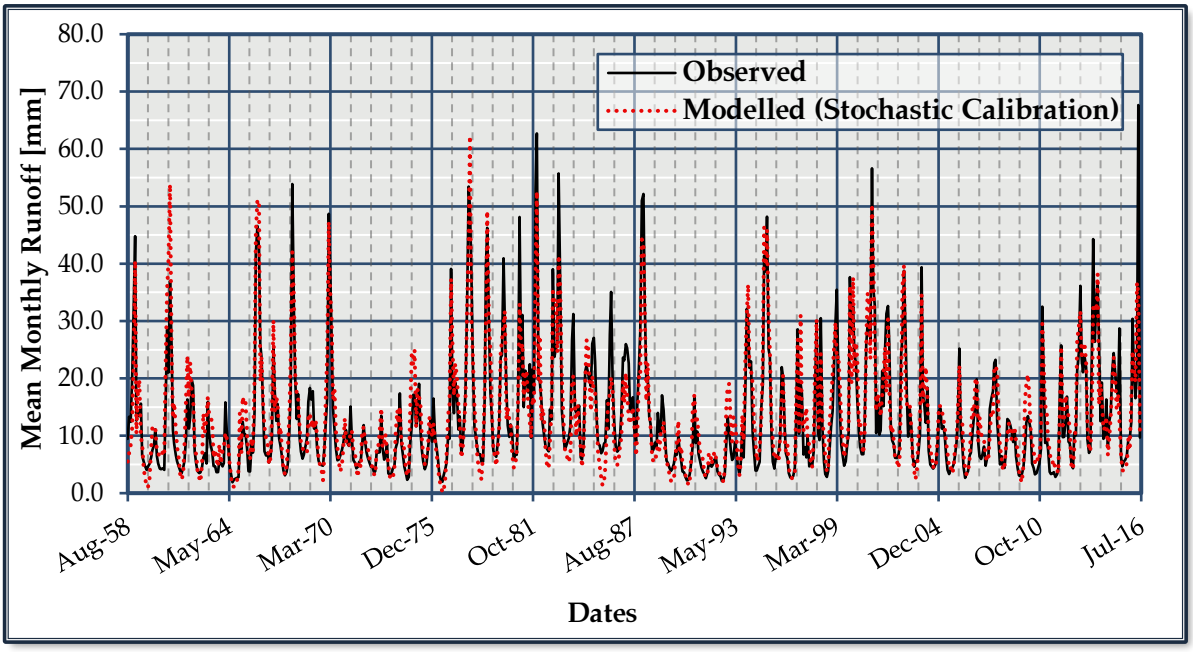


Figure 5.8 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach – **Zygos6P**)

As far it concerns the respective analysis with GR2M hydrological model, the results are given in **Table 5.3**. Once more, the results for the two calibration approaches converge. **Figure 5.9** and **Figure 5.10** represent the fitting of simulated values with each approach to historical data.

Table 5.3 | Results summary for GR2M hydrological model calibration, employing the split-sample approach and stochastic calibration at a monthly scale

	Model performance evaluation			**Optimized GR2M parameter values**	
Split-Sample Approach	**NSE_{Cal}**	**NSE_{Val}**	**NSE_{Tot}**	**X_1 [mm]**	**X_2**
	0.817	0.776	0.798	395.4	0.77
Stochastic Calibration	**NSE**			**X_1 [mm]**	**X_2**
	0.798			399.4	0.77
Abbreviations	**NSE_{Cal}**: Nash-Sutcliffe efficiency metric in **calibration** period **NSE_{Val}**: Nash-Sutcliffe efficiency metric in **validation** period **NSE_{Tot}**: overall Nash-Sutcliffe efficiency metric **NSE**: overall Nash-Sutcliffe efficiency metric				

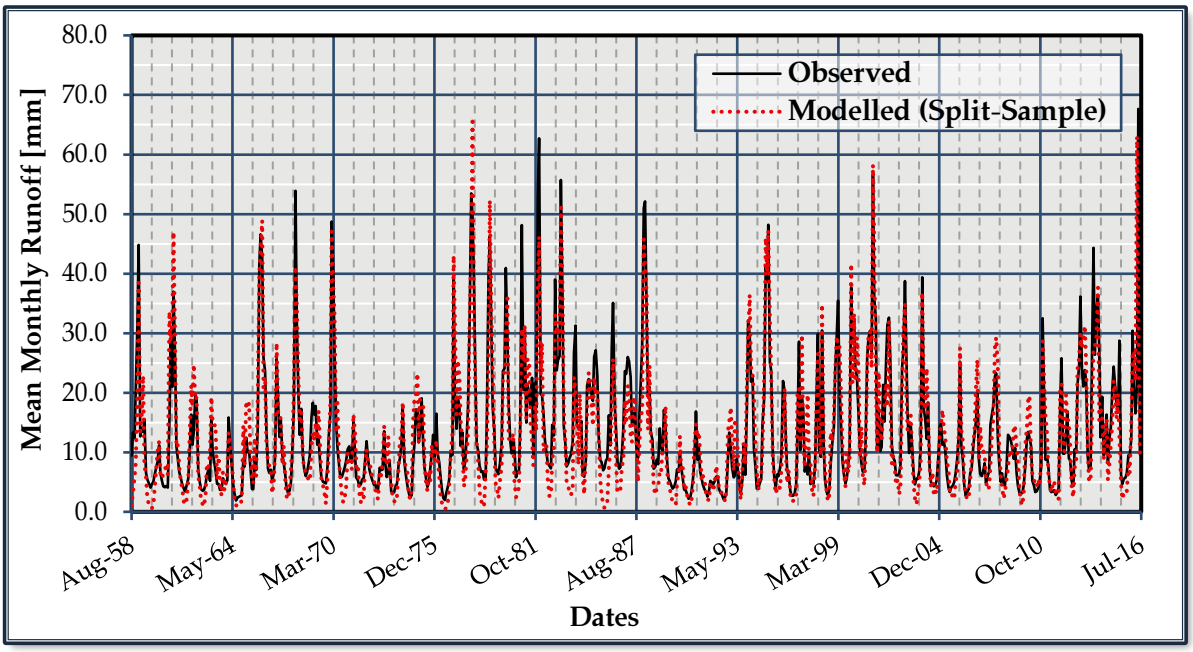


Figure 5.9 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach – **GR2M**)

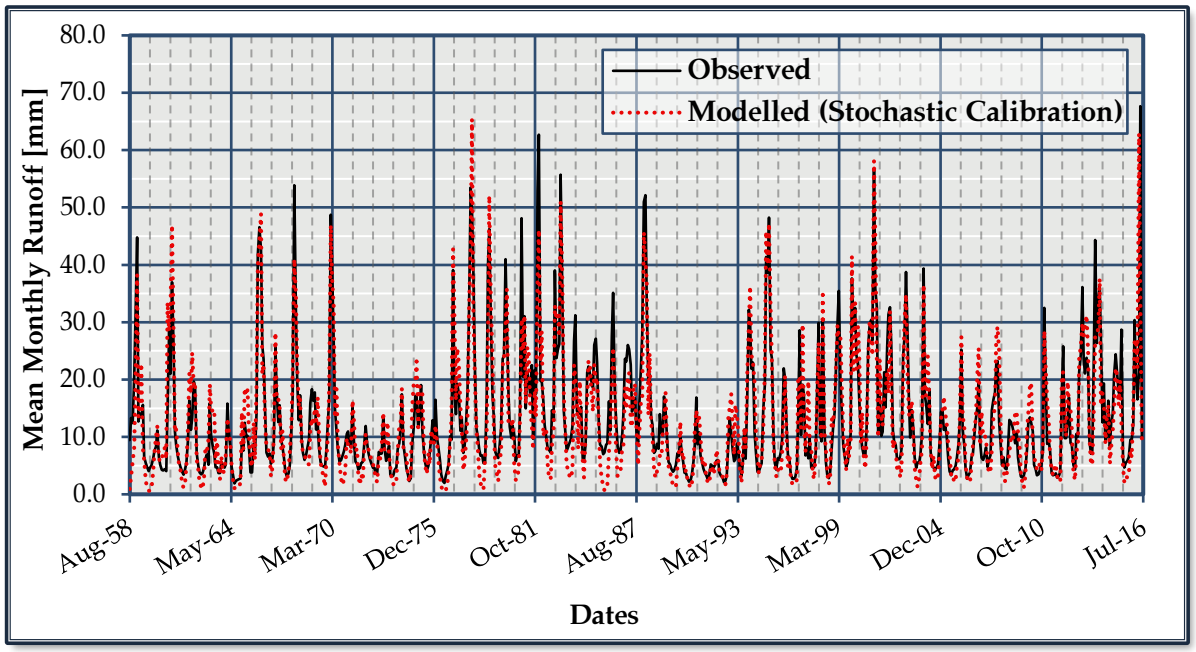


Figure 5.10 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach – **GR2M**)

5.3 Case Study B: Daily Time Scale Analysis

To demonstrate the potential of the stochastic calibration scheme at a daily scale, the daily observed data of the Loing catchment are employed (**Figure 5.2**). Because of the fine time scale, the hydrological models GR4J and GR6J are preferred for this analysis.

As in the monthly scale analysis, the split-sample scheme was initially applied, considering the first 29 years of historical records for calibration and the remaining 29 years for validation.

The generation of synthetic data of length 1000 years for the daily rainfall and runoff occurred again in accordance with the requirements discussed in section **3.2**, whereas the evapotranspiration input data for each model were estimated as the mean daily values of the parent data.

Subsequently, the selected hydrological models are calibrated against these modelled processes, and their structure is then validated against the full historical records (58 years).

The results of the daily scale analysis with GR4J model are given in **Table 5.4**. Contrasting model efficiency for the two calibration schemes, the split-sample procedure leads to better results, whilst there is an evident divergence between optimized parameters for each case. The performance of the model from the application of each calibration methodology can also be noticed in **Figure 5.11** and **Figure 5.12**.

Table 5.4 | Results summary for GR4J hydrological model calibration, employing the split-sample approach and stochastic calibration at a monthly scale

	Model performance evaluation			Optimized GR4J parameter values			
Split-Sample Approach	**NSE_{Cal}**	**NSE_{Val}**	**NSE_{Tot}**	**X_1 [mm]**	**X_2 [mm/d]**	**X_3 [mm]**	**X_4 [days]**
	0.848	0.895	0.871	520.1	-0.70	35.9	4.20
Stochastic Calibration	**NSE**			**X_1 [mm]**	**X_2 [mm/d]**	**X_3 [mm]**	**X_4 [days]**
	0.805			1011.6	-0.48	38.8	4.39
Abbreviations	**NSE_{Cal}**: Nash-Sutcliffe efficiency metric in **calibration** period **NSE_{Val}**: Nash-Sutcliffe efficiency metric in **validation** period **NSE_{Tot}**: overall Nash-Sutcliffe efficiency metric **NSE**: overall Nash-Sutcliffe efficiency metric						

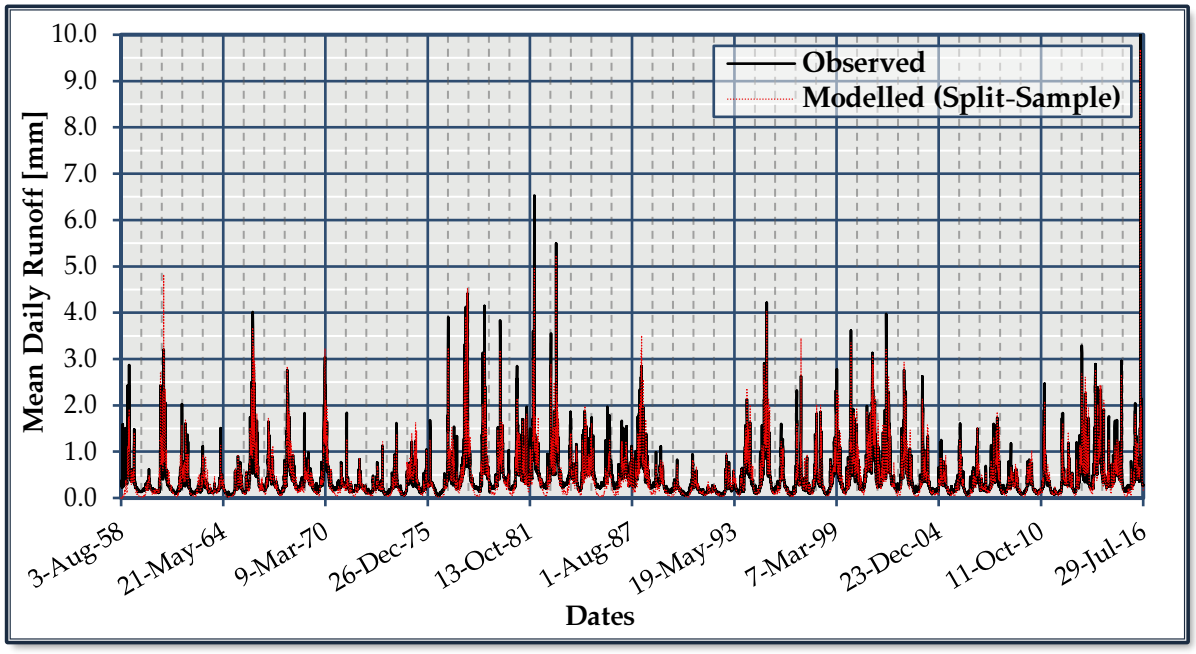


Figure 5.11 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach – **GR4J**)

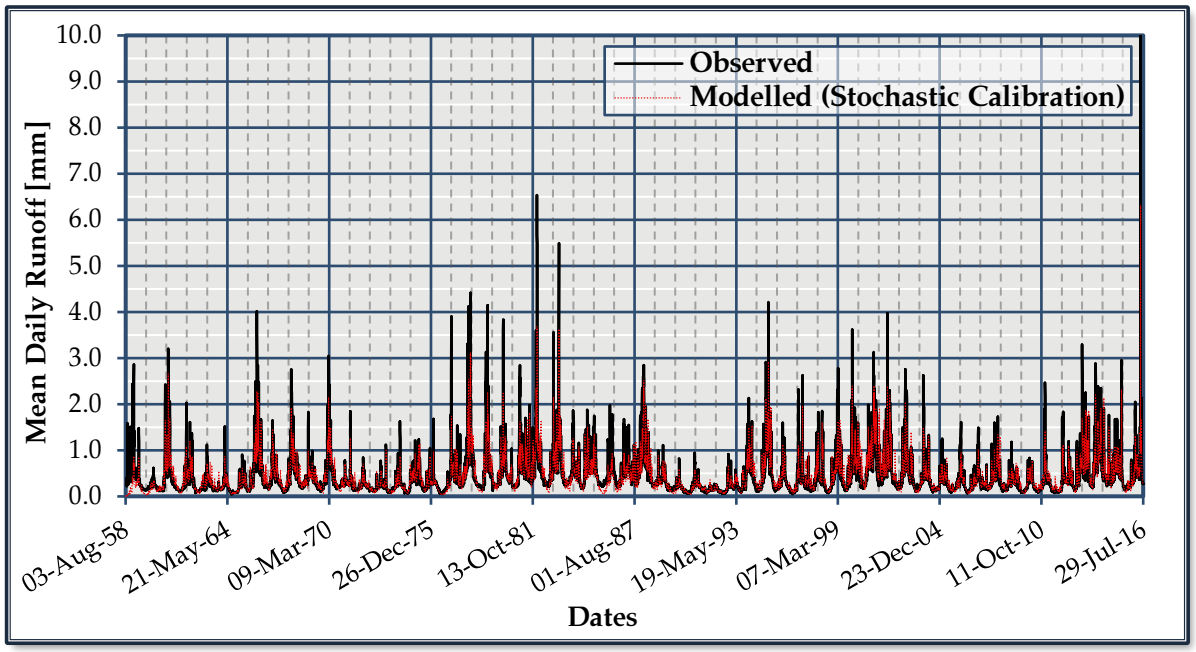


Figure 5.12 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach – **GR4J**)

From the corresponding analysis with GR6J hydrological model emerged results (**Table 5.5**), which denote that the conventional approach achieves a better performance, but then its outperformance is significantly restricted, rendering calibration with stochastic data still competent. **Figure 5.13** and **Figure 5.14** contrast historical data against simulated ones for each calibration scheme.

Table 5.5 | Results summary for GR6J hydrological model calibration, employing the split-sample approach and stochastic calibration at a monthly scale

	Model performance evaluation			**Optimized GR6J parameter values**					
Split-Sample Approach	**NSE_{Cal}**	**NSE_{Val}**	**NSE_{Tot}**	***X_1*** **[mm]**	***X_2*** **[mm/d]**	***X_3*** **[mm]**	***X_4*** **[days]**	***X_5*** **[-]**	***X_6*** **[mm]**
	0.849	0.802	0.858	242.1	-1.63	885.4	4.40	0.22	1.49
Stochastic Calibration	**NSE**			***X_1*** **[mm]**	***X_2*** **[mm/d]**	***X_3*** **[mm]**	***X_4*** **[days]**	***X_5*** **[-]**	***X_6*** **[mm]**
	0.815			296.6	-2.38	52.1	4.38	0.41	3.04
Abbreviations	**NSE_{Cal}**: Nash-Sutcliffe efficiency metric in **calibration** period **NSE_{Val}**: Nash-Sutcliffe efficiency metric in **validation** period **NSE_{Tot}**: overall Nash-Sutcliffe efficiency metric **NSE**: overall Nash-Sutcliffe efficiency metric								

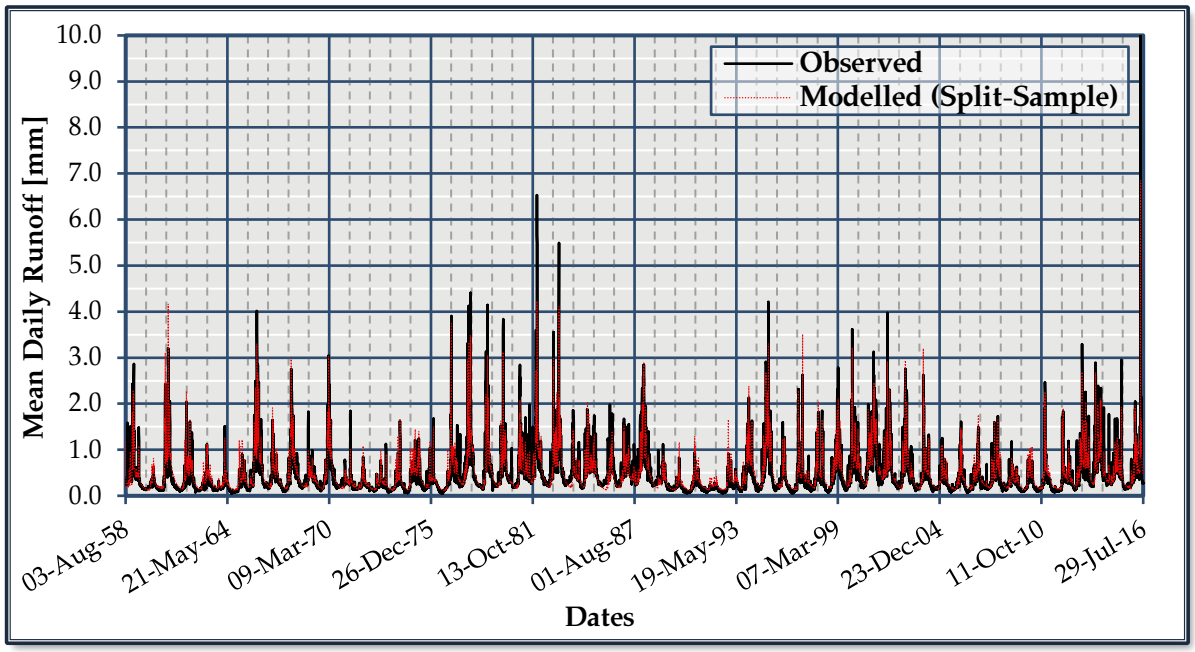


Figure 5.13 | Visual inspection of the agreement between observations and model predictions (**Split-Sample** approach – **GR6J**)

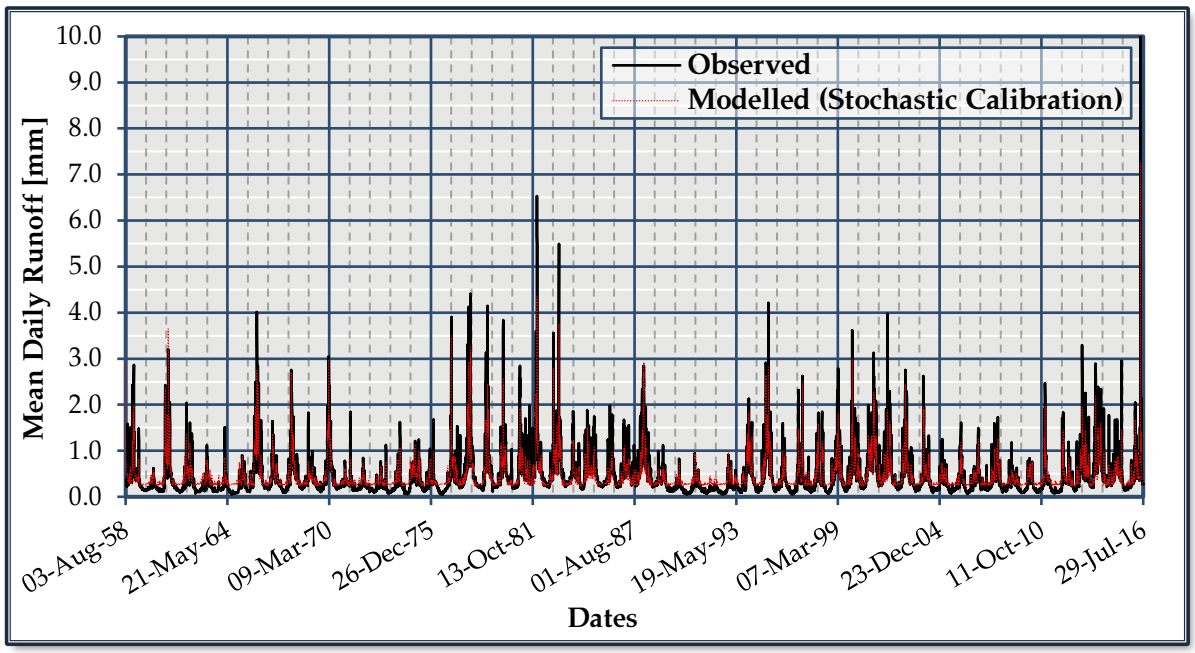


Figure 5.14 | Visual inspection of the agreement between observations and model predictions (**Stochastic Calibration** approach – **GR6J**)

5.4 Summary

In regard to the monthly time scale analysis, it is evident that the proposed methodology for calibration against synthetic rainfall-runoff data is functional. Specifically, the Zygos6P and GR2M hydrological models performed equally well for calibration against historical

data (split-sample approach) and against synthetic data (stochastic calibration approach). Moreover, the model parameter values for each calibration method are in agreement.

Concerning the daily time scale analysis, for the application of the split-sample scheme the overall efficiency metric (NSE_{Tot}) for the two daily hydrological models (GR4J and GR6J) is slightly higher than the respective efficiency metric (NSE) that was estimated for the stochastic calibration case. Hence, the convergence between the above-mentioned metrics indicate that the stochastic calibration framework has certain potentials for application also to the daily scale. On the other hand, it is also worth noticing that the optimized parameter values against the synthetic rainfall-runoff data are quite different with respect to the ones derived by calibrating against the half of historical data, especially for the case of GR6J model.

In addition, the results of this initial investigation constitute some first evidence to assume that the implementation of the stochastic calibration framework is independent of the chosen hydrological model.

6

LARGE-SCALE ANALYSIS

The next research step regards the implementation of the proposed framework for hydrological calibration in a large set of catchments at a monthly scale. Specifically, this study was conducted by selecting 100 catchments from the MOPEX database, accompanied by their respective datasets of hydrological processes. The location of each watershed across the USA is visible in **Figure 6.1**.

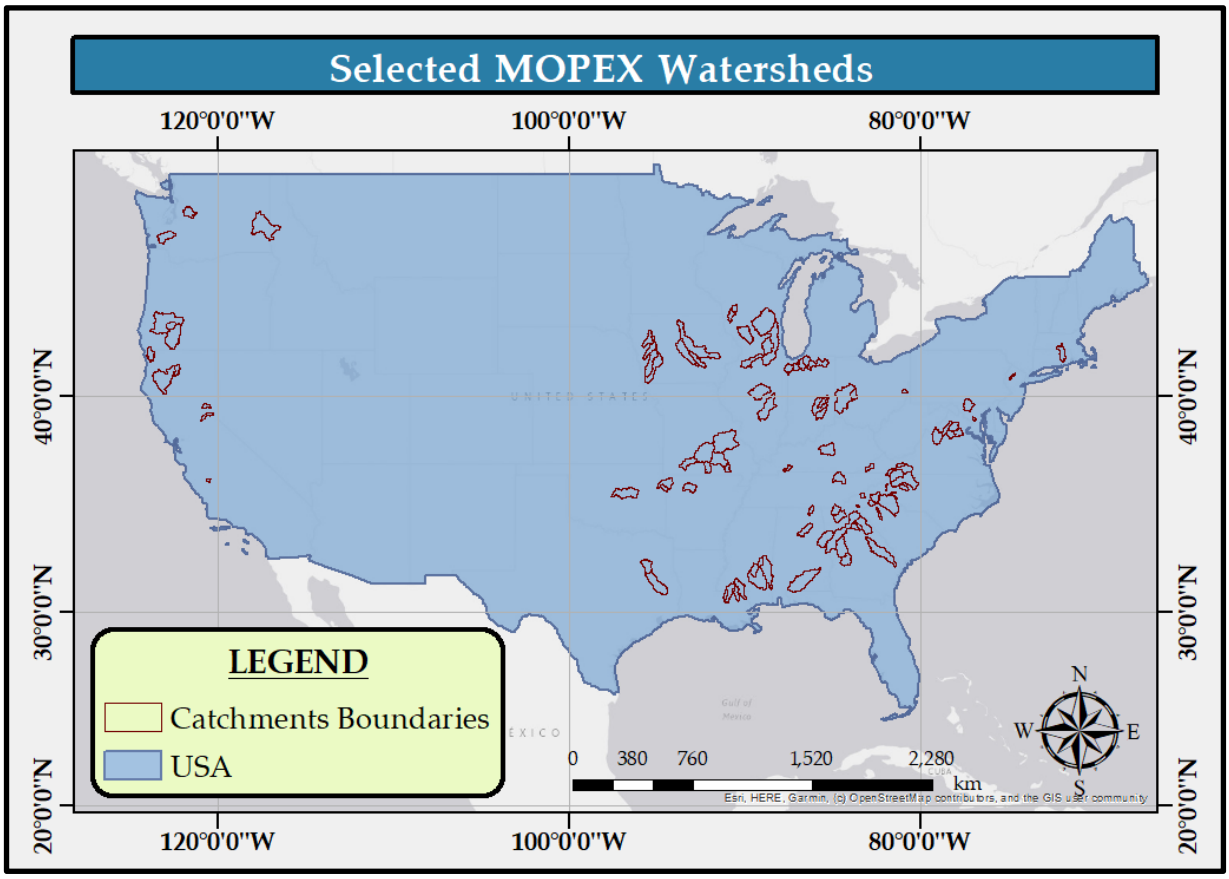


Figure 6.1 | Geographical location of the 100 selected MOPEX watersheds

The MOPEX (MOdel Parameter Estimation EXperiment) is a project developed for the enhancement of *a priori* parameter estimation methodologies for hydrological models and land surface parameterization schemes [Schaake et al., 2006].

For the selection of the optimal set of 100 catchments for this large-scale experiment, a combination of criteria was taken into account. Specifically, the selection criteria were:

- the percentage of stream flow missing values and
- cross-correlation between rainfall and streamflow.

6.1 Analysis Setup

The spatially averaged rainfall, evapotranspiration and streamflow data are provided for each of the 100 catchments at the daily time interval, thus the processes are aggregated at the monthly scale.

As in the previous research, several different hydrological models were used, to further fortify the assumption that hydrological calibration with stochastic inputs is independent of the chosen rainfall-runoff model structure. Specifically, for this study GR2M, Zygos4P and Zygos6P models were used, the structure of which has already been presented in sections **4.1 – 4.3**. It should be noted that also for this analysis, for the case of the split-sample approach, half of the total historical sample is used for the calibration procedure.

Aiming towards an efficient comparison between model performance for each calibration methodology, it was decided to present graphically the model efficiency (NSE) over the period of the historical sample, which is allocated for validation purposes. This decision stemmed from the notion that the validation period of observed records contains the only data over which, in both calibration methodologies, the rainfall-runoff model has not been trained. On the contrary, in the case of the split-sample calibration approach, model efficiency over the calibration period of historical data is biased, since the selected model has been over-fitted on this period.

6.2 GR2M Model Analysis

The results for model calibration with this two-parameter hydrological model indicate that in 62% of cases model calibration with synthetic data outperformed the one which was implemented with a real-world calibration dataset. This conclusion is also reflected in **Figure 6.2**. For the sake of completeness, it should be pointed out that, when inspecting the performance of GR2M model over the calibration period of the historical sample, the case of model training with simulated data managed to do better just for one catchment, which can be attributed to the model being over-fitted on the calibration subset of the observations. As far it concerns the overall model efficiency, it was estimated that for 42 out of the 100 examined catchments the stochastic calibration methodology achieved better results than the split-sample approach.

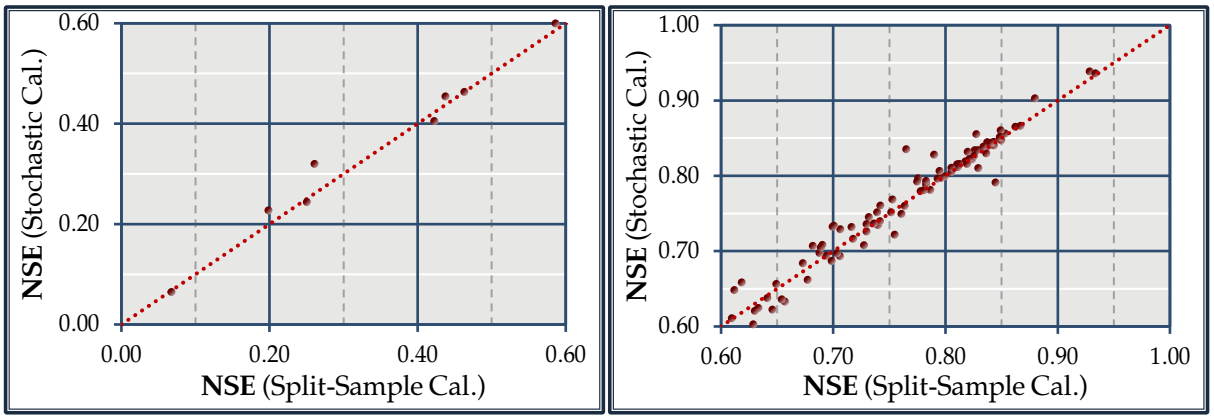


Figure 6.2 | Scatter plots of poor/low (left) or good/high (right) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration (GR2M model)

The scatter plot for model parameter X_1 (**Figure 6.3)** depicts that the adjusted values inferred from the stochastic data are substantially differentiated from the data-driven ones, whereas in the case of model parameter X_2 there is a significant convergence between the values emerged from the two procedures.

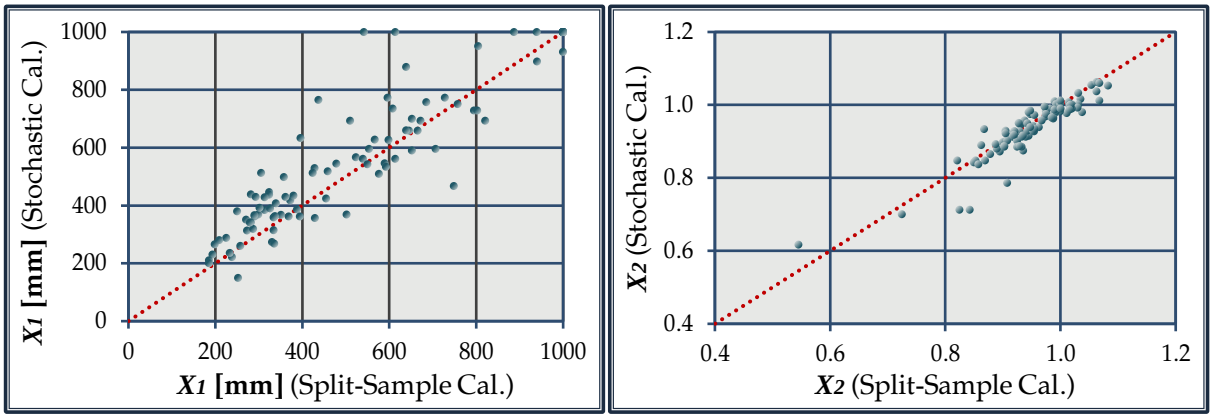


Figure 6.3 | Scatter plots of GR2M model parameters estimated through Split-Sample approach and Stochastic Calibration

6.3 Zygos4P Model Analysis

In the case of the Zygos4P model, the calibration approach with synthetic data ensured a better model performance for a significant proportion (67%) of the basins, in contrast to the calibration method that the model is fed with historical inputs (**Figure 6.4**).

Furthermore, **Figure 6.5** demonstrates that the optimized K, κ and λ parameters inferred from the stochastic data and the respective ones derived from the classical method present a strong divergence pattern, while the ν parameter values for the two methodologies are generally in agreement.

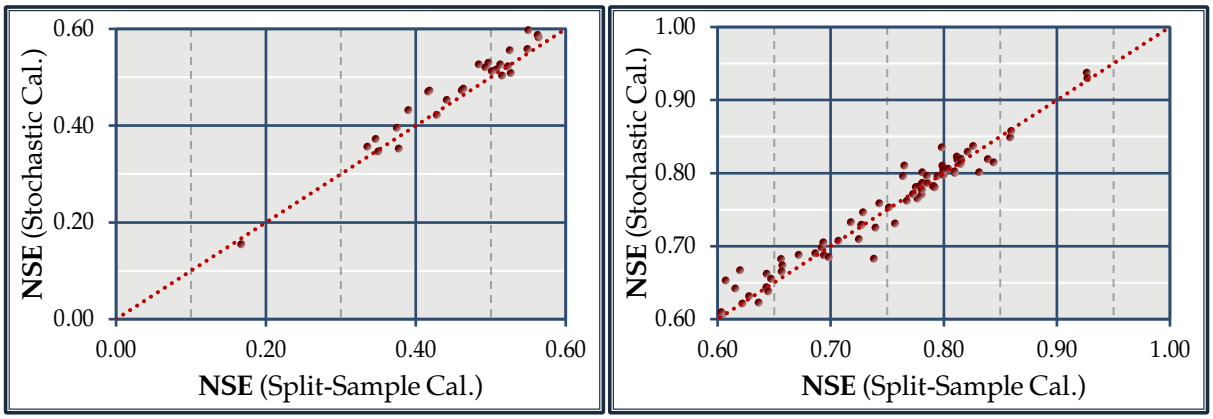


Figure 6.4 | Scatter plots of poor/low (left) or good/high (right) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration (Zygos4P model)

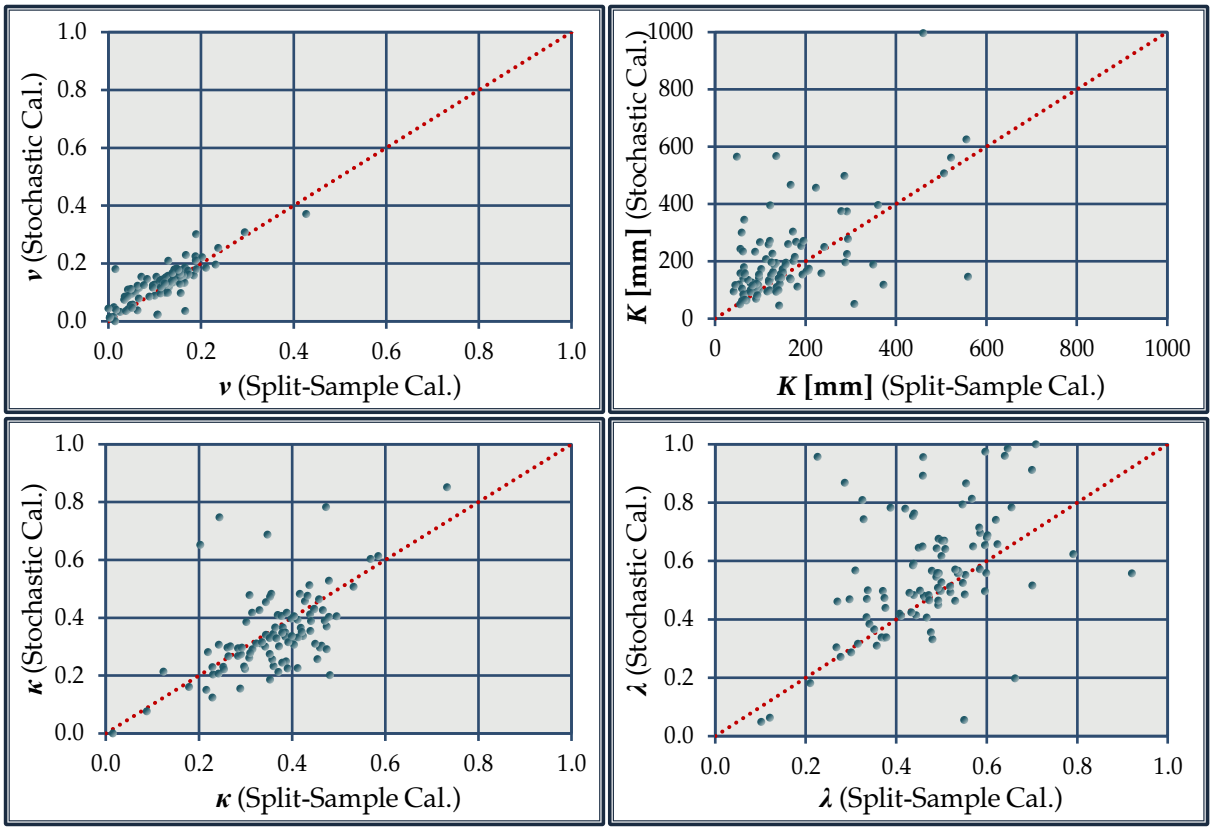


Figure 6.5 | Scatter plots of Zygos4P model parameters estimated through Split-Sample approach and Stochastic Calibration

6.4 Zygos6P Model Analysis

In order to further benchmark the performance of the proposed framework, another investigation was conducted, this time by using the six-parameter model Zygos6P. The results of the analysis are given graphically in **Figure 6.6** and it can be concluded that the calibration scheme with stochastic data achieves a better approximation (60%) of the hydrologic response of each basin than the conventional calibration approach.

Moreover, from the presented scatter plots in **Figure 6.7** it can be noticed that there is a strong disperse of the points around the bisector of each plot, which denotes that, overall, the adjusted parameter values according to the split-sample approach do not agree with the corresponding values resulted from stochastic calibration.

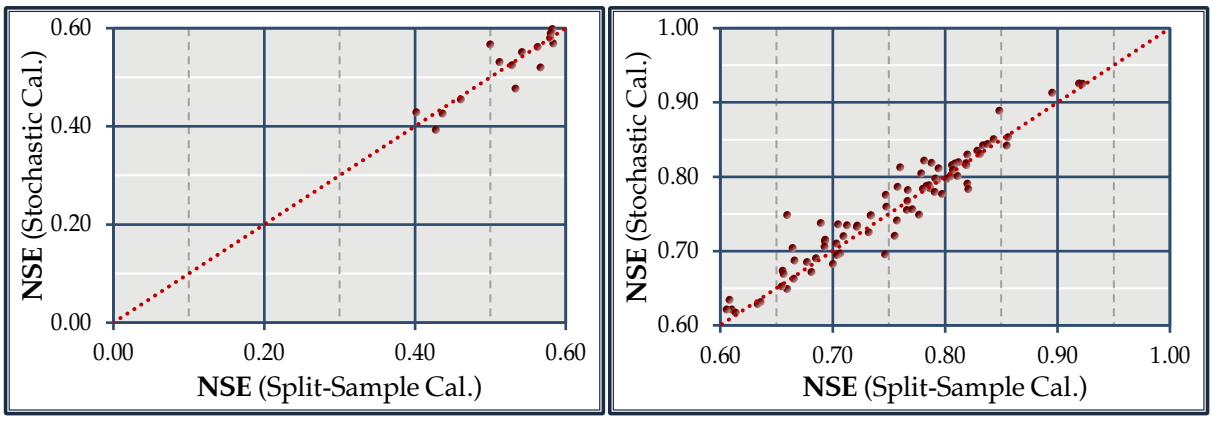
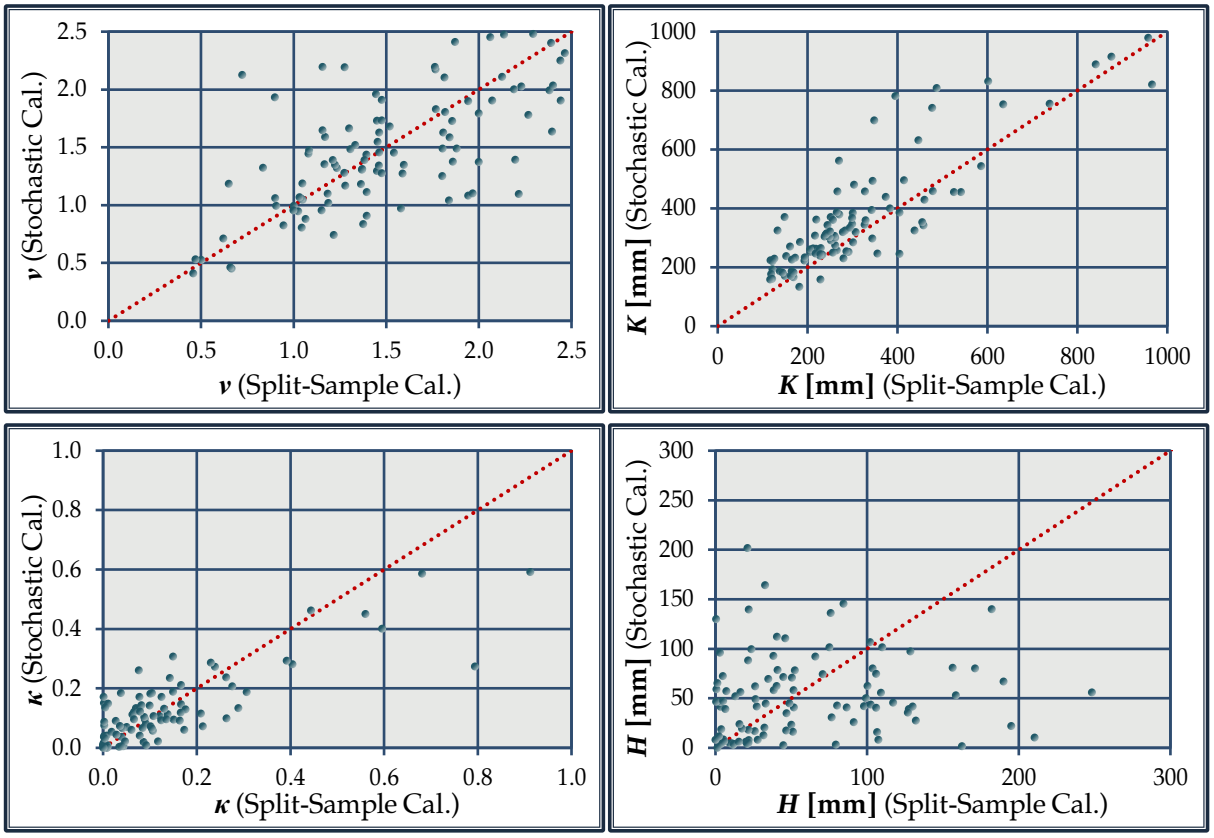


Figure 6.6 | Scatter plots of poor/low (left) or good/high (right) model performance (NSE) by employing the Split-Sample approach and the Stochastic Calibration (Zygos6P model)



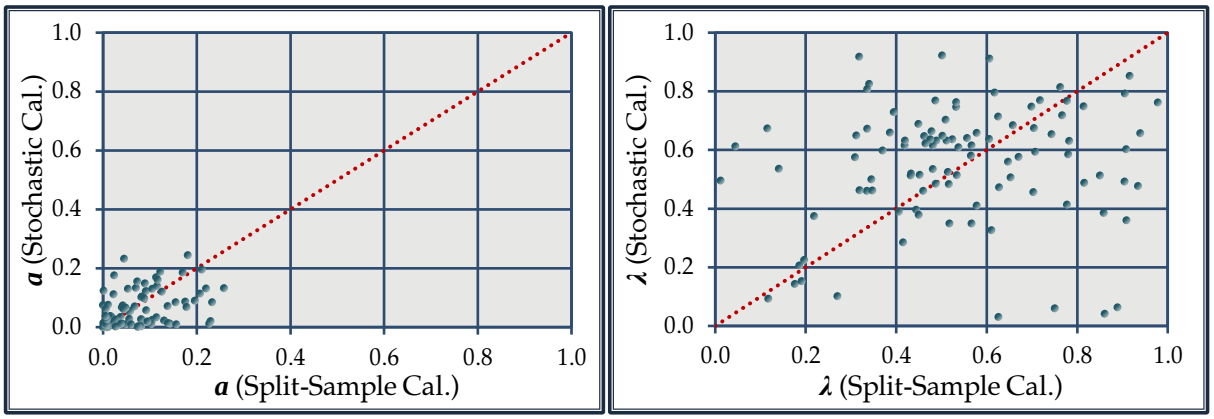


Figure 6.7 | Scatter plots of Zygos6P model parameters estimated through Split-Sample approach and Stochastic Calibration

6.5 Conclusions

After reviewing the results of 100 different test-cases and while keeping in mind the proposed comparison aspect between the two tested calibration methodologies, it seems clear that the stochastic calibration outperforms the conventional split-sample approach for each model case, as the former presents better NSE index values for the majority of the examined catchments. Hence, this evidence emphasizes the applicability of the proposed method, in spite of the model complexity.

Additionally, it is evident from each model analysis that some parameter values estimated by the stochastic calibration procedure differ substantially from the ones stemmed from the data-driven approach. Concerning the GR2M model analysis, values for X_1 parameter vary significantly for the split-sample calibration approach and for the stochastic simulation approach. Regarding the analysis using the Zygos4P model, parameter values for K, κ and λ, derived from the use of the two calibration methodologies, disagree in most of the tested cases, with this divergence being stronger in the case of parameter λ. Finally, discrepancies between the parameter values, estimated by each calibration methodology, appear in each parameter in the case of the Zygos6P model.

However, it worth mentioning that during this large-scale experiment, the use of the long stochastic inputs in conjunction with the optimization algorithms, used for the calibration process, pose a serious barrier in the application of the proposed framework, in terms of computational effort.

7

DISCUSSION AND SUGGESTIONS FOR FUTURE RESEARCH

7.1 Summary and conclusive remarks

The purpose of this research was to introduce a stochastic-simulation framework for hydrological calibration, as well as to present its strengths against conventional split-sample approaches.

Initially, a monthly-scale analysis was employed by means of proof-of-concept, and the results highlighted that this novel methodology proved equally sufficient as the classical split-sample scheme in terms of model performance, for both of the two chosen hydrological models, with the resulting parameters from each method and model being in agreement.

On the other hand, the daily-scale analysis, which was performed for the same case study and for a four-parameter and six-parameter model, demonstrated that the split-sample scheme performed slightly better than the stochastic calibration, though without invalidating the potentials of the method at this time scale. However, it should be noted that there is a noticeable difference between the model parameter values derived from each calibration approach.

Next research steps occurred in the direction of a large-scale experiment, involving a set of 100 catchments of the MOPEX database and three hydrological models of varying complexity. The results denote that for most of the tested cases the stochastic calibration framework is capable of outperforming the split-sample approach, regardless of the chosen model for hydrological simulation, therefore rendering calibration with the use of synthetic inputs and outputs independent of the hydrological model complexity.

Regarding the divergence between the estimated parameter values, stemmed from each calibration methodology, it proved to be more remarkable as the tested hydrological model structure progressed to be more complex.

7.2 Proposals for Future Research

Undoubtedly, this is just a first attempt to align efficient hydrological calibration with stochastics, and calls for further studies. More specifically:

- It may not always be efficient to characterize the different aspects of model

performance for a particular rainfall-runoff model with only one performance metric. Thus, more criteria for model evaluation should be compared, in order to evaluate the predictive capacity of each method.

- It is also worth exploring the potentials of stochastic calibration by employing rainfall-runoff models of more complex structure, since more parsimonious mode structures were selected for this investigation.
- Another suggestion would be to further test the applicability of the approach explored in the present research at the daily time scale, by employing an extensive benchmarking experiment, such as the one that was undertaken in this study for the monthly time scale.
- As mentioned before, a challenging issue is the significantly time-consuming computational process in the case of a large-scale analysis, due to the use of substantially long synthetic data (length of thousands of years). A combination of the hereby presented methodologies with state-of-the-art global optimization methods (e.g., [Tsoukalas *et al.*, 2016]), such as machine learning and surrogate-based optimization techniques, will yield interesting results, as regards computational effort.

REFERENCES

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., . . . Jha, M. K. (2012). SWAT: model use, calibration, and validation. *Transactions of the ASABE, 55*(4), 1491-1508. doi:10.13031/2013.42256

Arsenault, R., Brissette, F., & Martel, J. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology, 566*, 346-362. doi:10.1016/j.jhydrol.2018.09.027

Arsenault, R., Essou, G. R., & Brissette, F. P. (2017). Improving hydrological model simulations with combined multi-input and multimodel averaging frameworks. *Journal of Hydrologic Engineering, 22* (4), 04016066. doi:10.1061/(ASCE)HE.1943-5584.0001489

Brigode, P., Oudin, L., & Perrin, C. (2013). Hydrological model parameter instability: a source of additional uncertainty in estimating the hydrological impacts of climate change? *Journal of Hydrology, 476*, 410–425. doi:10.1016/j.jhydrol.2012.11.012

Coron, L., Andréassian , V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions :. *Water Resources Research, 48* (5), 17. doi:10.1029/2011WR011721

Coron, L., Andréassian, V., Bourqui, M., Perrin, C., & Hendrickx, F. (2011). Pathologies of hydrological models used in changing climatic conditions: a review. In *S.W. Franks, et Al., Eds.* (Vol. 344, pp. 39–44). IAHS Publication.

Coron, L., Andréassian, V., Perrin, C., Bourqui, M., & Hendrickx, F. (2014). On the lack of robustness of hydrologic models regarding water balance simulation: A diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. *Hydrol. Earth Syst. Sci., 18* (2), 727–746. doi:10.5194/hess-18-727-2014

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., & Michel, C. (2021). *airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling*, R package version 1.6.10.4. doi:10.15454/EX11NA

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling and Software, 94*, 166-171. doi:10.1016/j.envsoft.2017.05.002

Duethmann, D., Blöschl, G., & Parajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrol. Earth Syst. Sci., 24*, 3493–3511. doi:10.5194/hess-24-3493-2020

Efstratiadis, A., & Koutsoyiannis, D. (2002). An evolutionary annealing-simplex algorithm for global optimisation of water resource systems. *Proceedings of the Fifth International Conference on Hydroinformatics* (pp. 1423 – 1428). Cardiff, UK: International Water Association.

Efstratiadis, A., Dialynas, Y., Kozanis, S., & Koutsoyiannis, D. (2014). A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environmental Modelling and Software, 62*, 139–152. doi:10.1016/j.envsoft.2014.08.017

Efstratiadis, A., Tsoukalas, I., & Kossieris, P. (n.d.). Improving model identifiability by embedding stochastic simulation within hydrological calibration. In D. Solomatine, & G. Corzo (Eds.), *Advances in Hydroinformatics: Artificial Intelligence and Optimization for Water Resources* (Vol. Edited volume of AGU Books / Wiley (in review)).

Essou, G. R., Arsenault, R., & Brissette, F. P. (2016). Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology, 537*, 334-345. doi:10.1016/j.jhydrol.2016.03.063

Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. H. (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences, 17* (1), 149–161. doi:10.5194/hess-17-149-2013

Gowda, P. H., Mulla, D. J., Desmond, E. D., Ward, A. D., & Moriasi, D. N. (2012). ADAPT: Model use, calibration, and validation. *Transactions of the ASABE, 55* (4), 1345-1352. doi:10.13031/2013.42246

Gupta, V. K., & Sorooshian, S. (1985). The relationship between data and the precision of parameter estimates of hydrological models. *Journal of Hydrology, 81* (1-2), 57-77. doi:10.1016/0022-1694(85)90167-2

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal, 31* (1), 13–24. doi:10.1080/02626668609491024

Kossieris, P., Tsoukalas, I., Makropoulos, C., & Savic, D. (2019). Simulating marginal and dependence behaviour of water demand processes at any fine time scale. *Water, 11*(5), 885. doi:10.3390/w11050885

Koutsoyiannis, D. (2013). Hydrology and Change. *Hydrological Sciences Journal, 58* (6), 1177–1197. doi:10.1080/02626667.2013.804626

Koutsoyiannis, D. (2020). Simple stochastic simulation of time irreversible and reversible processes. *Hydrological Sciences Journal, 65* (4), 536–551. doi:10.1080/02626667.2019.1705302

Kozanis, S., & Efstratiadis, A. (2006). Zygos: A basin processes simulation model. *21st European Conference for ESRI Users.* Athens, Greece.

Le Moine, N. (2008). *Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modéles pluie-débit?* (Doctoral dissertation, Doctorat Géosciences et Ressources Naturelles, Université Pierre et Marie Curie Paris VI).

Matalas, N. C., & Wallis, J. R. (1976). *Generation of synthetic flow sequences, Systems Approach to Water Management.* (A. K. Biswas, Ed.) New York: McGraw-Hill.

Michel, C. (1991). Hydrologie appliquée aux petits bassins ruraux. *Hydrology Handbook (in French)*.

Mouelhi, S., Michel, C., Perrin, C., & Andréassian, V. (2006). Stepwise Development of a Two-Parameter Monthly Water Balance Model. *Journal of Hydrology, 318* (1 - 4), 200 - 214. doi:10.1016/j.jhydrol.2005.06.014

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part 1: a discussion of principles. *Journal of Hydrology, 10* (3), 282 – 290. doi:10.1016/ 0022-1694(70)90255-6

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a Parsimonious Model for Streamflow Simulation. *Journal of Hydrology, 279*, 275 - 289. doi:10.1016/S0022-1694(03)00225-7

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., & Andréassian, V. (2011). A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. *Journal of Hydrology, 411* (1–2), 66 - 76. doi:10.1016/j.jhydrol.2011.09.034.

Razavi, S., & Tolson, B. A. (2012). An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research, 49*, 8418-8431. doi:10.1002/2012WR013442

Razavi, S., & Tolson, B. A. (2013). An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research, 49*, 8418-8431. doi:10.1002/2012WR013442

Rebolho, C., Andréassian, V., & Le Moine, N. (2018). Inundation mapping based on reach-scale effective geometry. *Hydrology and Earth System Sciences, 22* (11), 5967-5985. doi:10.5194/hess-22-5967-2018

Rebolho, C., Andréassian, V., Tsoukalas, I., & Efstratiadis, A. (2018). La crue du Loing de Juin 2016 était-elle exceptionnelle? *De la prévision des crues à la gestion de crise.* Avignon: Société Hydrotechnique de France.

Refsgaard, J. C. (2000). Towards a formal approach to calibration and validation of models using spatial data. In *Spatial Patterns in Catchment Hydrology: Observations and Modeling* (pp. 329-354). Cambridge, U.K.: Cambridge University Press.

Schaake, J., Cong, S., & Duan, Q. (2006). The US MOPEX data set. *Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment – MOPEX, 307*, 9-28.

Seibert, J., & McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research, 38* (11), 23-1-23-14. doi:10.1029/2001WR000978

Singh, S. L., & Bárdossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Advances in Water Resources, 38*, 81-91. doi:10.1016/j.advwatres.2011.12.006

Thirel, G., Andréassian, V., Perrin, C., Audouy, N.-L., Berthet, L., Edwards, P., . . . Vaze, J. (2015b). Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal, 60* (7-8), 1184-1199. doi:10.1080/02626667.2014.967248

Tsoukalas, I., Efstratiadis, A., & Makropoulos, C. (2018a). Stochastic periodic autoregressive to anything (SPARTA): Modelling and simulation of cyclostationary processes with

arbitrary marginal distributions. *Water Resources Research, 54* (1), 161–185. doi:10.1002/2017WR021394

Tsoukalas, I., Efstratiadis, A., & Makropoulos, C. (2019). Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *Journal of Hydrology, 575*, 354–380. doi:10.1016/j.jhydrol.2019.05.017

Tsoukalas, I., Kossieris, P., & Makropoulos, C. (2020). Simulation of non-Gaussian correlated random variables, stochastic processes and random fields: Introducing the anySim R-package for environmental applications and beyond. *Water, 12* (6), 1645. doi:10.3390/w12061645

Tsoukalas, I., Kossieris, P., Efstratiadis, A., & Makropoulos, C. (2016). Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. *Environ. Model. Soft., 77*, 122-142. doi:10.1016/j.envsoft.2015.12.008

Tsoukalas, I., Papalexiou, S. M., Efstratiadis, A., & Makropoulos, C. (2018b). A cautionary note on the reproduction of dependencies through linear stochastic models with non-Gaussian white noise. *Water, 10* (6), 771. doi:10.3390/w10060771

APPENDICES

Appendix A: R script for "Zygos4P" and "Zygos6P" hydrological models

Appendix A1: *"Zygos4P" Model*

```
#############################################################################
#                  FUNCTION : Zygos4P Hydrological Model                    #
#############################################################################

Zygos4P = function(Pars , P , ET_pot , Q_obs, Cal_Per)  {

  # A monthly lumped hydrological model with 4 parameters

  ##################### INPUT DATA ####################
  # P           [mm] : Precipitation                 #
  # ET_pot      [mm] : Potential Evapotranspiration  #
  # Q_obs       [mm] : Observed Runoff               #
  # Pars             : Model Parameters              #
  # Cal_Per [months] : Chosen Calibration Period     #
  #####################################################

  ################## MODEL PARAMETERS & RANGES ###############
  # v       : Surface runoff coefficient        | 0 - 1      #
  # K [mm]  : Upper soil water storage capacity | 0 - 1000   #
  # k       : Groundwater recharge coefficient  | 0 - 1      #
  # lamda   : Groundwater runoff coefficient    | 0 - 1      #
  ############################################################

  v     = Pars[[1]]
  K     = Pars[[2]]
  k     = Pars[[3]]
  lamda = Pars[[4]]


  #### PREALOCATION OF ARRAYS ####

  # S_ini    [mm]  : Soil Moisture at the start of the current month
  # S_ini(1) [mm]  : Initial Soil Moisture, equal to K/5
  # G_ini    [mm]  : Groundwater Storage at the start of the current month
  # G_ini(1) [mm]  : Initial Groundwater Storage, equal to 200 mm
  # S_fin    [mm]  : Soil Moisture at the end of the current month
```

```
# G_fin    [mm]   : Groundwater Storage at the end of the current month
# Q_dir    [mm]   : Direct Runoff
# ET_dir   [mm]   : Direct Evapotranspiration
# INFIL    [mm]   : Infiltration
# ET_soil  [mm]   : Soil Evapotranspiration
# ET_act   [mm]   : Actual Evapotranspiration
# PERC     [mm]   : Percolation
# Q_soil   [mm]   : Upper Soil Runoff
# Q_base   [mm]   : Baseflow
# Q_sim    [mm]   : Simulated Runoff
# dQ      [mm^2]  : Squared Differences of Q_obs and Q_sim
# dQ_obs  [mm^2]  : Squared Differences of Q_obs and mean value of Q_obs

t_max = length(P)

S_ini    = matrix(data = 0, nrow = t_max, ncol = 1)
S_ini[1] = K / 5

G_ini    = rep(S_ini)
G_ini[1] = 0

S_fin    = G_fin  = Q_dir = ET_dir = INFIL =
ET_soil  = ET_act = PERC  = Q_soil = Q_base =
Q_sim    = dQ     = dQ_obs =    rep(S_ini)


#### MODEL ####

for ( t in 1:t_max ) {

  ### SOIL SURFACE PROCESSES ###
  Q_dir[t]   = v * P[t]
  ET_dir[t]  = min( (1-v) * P[t] , ET_pot[t] )
  INFIL[t]   = P[t] - Q_dir[t] - ET_dir[t]

  ### SOIL UNSATURATED ZONE PROCESSES ###
  S          = S_ini[t] + INFIL[t]
  ET_soil[t] = S * (1 - exp(- ( ET_pot[t] - ET_dir[t] ) / K ))
  S          = S - ET_soil[t]
  PERC[t]    = k * S
  S          = S - PERC[t]
  Q_soil[t]  = max( S - K , 0 )
  S_fin[t]   = S - Q_soil[t]
```

```
  ### SOIL SATURATED ZONE PROCESSES ###
  G          = G_ini[t] + PERC[t]
  Q_base[t]  = lamda * G
  G_fin[t]   = G - Q_base[t]

  if ( t == t_max ) { break }
  else  {
    S_ini[t+1] = S_fin[t]
    G_ini[t+1] = G_fin[t]
  }
}

ET_act = ET_dir + ET_soil
Q_sim  =  Q_dir +  Q_soil + Q_base


#### NASH - SUTCLIFFE EFFICIENCY (NSE) ####

### CALIBRATION PERIOD ###
for ( i in 1:Cal_Per ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[1:Cal_Per] ) )^2
}
NSE_Cal = 1 - sum( dQ[1:Cal_Per] ) / sum( dQ_obs[1:Cal_Per] )

### VALIDATION PERIOD ###
j = Cal_Per + 1
for ( i in j:length(P) ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[j:length(P)] ) )^2
}
NSE_Val = 1 - sum( dQ[j:length(P)] ) / sum( dQ_obs[j:length(P)] )

### WHOLE PERIOD ###
for ( i in 1:length(P) ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[1:length(P)] ) )^2
}
NSE_Tot = 1 - sum( dQ[1:length(P)] ) / sum( dQ_obs[1:length(P)])


#### MODEL OUTPUT ####

column.names = c('NSE_Calibration','NSE_Validation','NSE_Total')
```

```
  resultsTable = array(c(format(NSE_Cal, digits = 4, width = 11,
                                justify = 'centre'),
                         format(NSE_Val, digits = 4, width = 10,
                                justify = 'centre'),
                         format(NSE_Tot, digits = 4, width = 7,
                                justify = 'centre')),
                       dim = c(1,3), dimnames = list('Value',column.names))


  row.names    = seq(from = 1, to = t_max, by = 1)
  column.names = c('Simulated_Runoff','Actual_ET',
                   'Soil_Moisture','Groundwater_Storage')

  operationTS  = array(c(format(Q_sim,  digits = 3, width = 10,
                                justify = 'centre'),
                         format(ET_act, digits = 3, width = 7,
                                justify = 'centre'),
                         format(S_fin,  digits = 3, width = 11,
                                justify = 'centre'),
                         format(G_fin,  digits = 3, width = 14,
                                justify = 'centre')),
                       dim = c(t_max,4), dimnames = list(row.names,column.names))

  class(operationTS) = 'numeric'

  output = list(resultsTable , operationTS)

  print(resultsTable, quote = FALSE)

  return( output )

}
```

Appendix A2: *"Zygos6P" Model*

```
#############################################################################
#                  FUNCTION : Zygos6P Hydrological Model                    #
#############################################################################

Zygos6P = function(Pars , P , ET_pot , Q_obs, Cal_Per)  {

  # A monthly lumped hydrological model with 6 parameters
```

```
#################### INPUT DATA ####################
# P          [mm] : Precipitation                  #
# ET_pot      [mm] : Potential Evapotranspiration  #
# Q_obs       [mm] : Observed Runoff               #
# Pars             : Model Parameters              #
# Cal_Per [months] : Chosen Calibration Period     #
####################################################

################# MODEL PARAMETERS & RANGES #################
# v      : Surface runoff coefficient         |    1 - 2.5  #
# K [mm] : Upper soil water storage capacity  |    0 - 1000 #
# k      : Groundwater recharge coefficient   |    0 - 1.0  #
# H [mm] : Threshold in the 2nd reservoir     |             #
#           above which baseflow occurs       |    0 - 300  #
# a      : Losses coefficient (vertical       |             #
#           downflow from the 2nd reservoir)  |    0 - 1.0  #
# lamda  : Baseflow coefficient (horizontal   |             #
#           downflow from the 2nd reservoir)  | 0.01 - 1.00 #
#############################################################

v     = Pars[[1]]
K     = Pars[[2]]
k     = Pars[[3]]
H     = Pars[[4]]
a     = Pars[[5]]
lamda = Pars[[6]]


#### PREALOCATION OF ARRAYS ####

# S_ini    [mm]  : Soil Moisture at the start of the current month
# S_ini(1) [mm]  : Initial Soil Moisture, equal to K/5
# G_ini    [mm]  : Groundwater Storage at the start of the current month
# G_ini(1) [mm]  : Initial Groundwater Storage, equal to 1.1 * H
# S_fin    [mm]  : Soil Moisture at the end of the current month
# G_fin    [mm]  : Groundwater Storage at the end of the current month
# Q_dir    [mm]  : Direct Runoff
# ET_dir   [mm]  : Direct Evapotranspiration
# INFIL    [mm]  : Infiltration
# ET_soil  [mm]  : Soil Evapotranspiration
# ET_act   [mm]  : Actual Evapotranspiration
# PERC     [mm]  : Percolation
# Q_soil   [mm]  : Upper Soil Runoff
```

```
  # Q_base   [mm]   : Baseflow
  # Q_sim    [mm]   : Simulated Runoff
  # dQ       [mm^2] : Squared Differences of Q_obs and Q_sim
  # dQ_obs   [mm^2] : Squared Differences of Q_obs and mean value of Q_obs

t_max = length(P)

S_ini    = matrix(data = 0, nrow = t_max, ncol = 1)
S_ini[1] = K / 5

G_ini    = rep(S_ini)
G_ini[1] = 1.1 * H

S_fin    = G_fin  = Q_dir = ET_dir = INFIL =
ET_soil  = ET_act = PERC  = Q_soil = Loss  =
Q_base   = Q_sim  = dQ    = dQ_obs = rep(S_ini)


#### MODEL ####

  for ( t in 1:t_max ) {

    ### SOIL SURFACE PROCESSES ###
    ET_dir[t]  = min( 0.5 * P[t] , ET_pot[t] )
    Q_dir[t]   = ( P[t] - ET_dir[t] ) * ( S_ini[t] / K )^v
    INFIL[t]   = P[t] - ET_dir[t] - Q_dir[t]

    ### SOIL UNSATURATED ZONE PROCESSES ###
    S          = S_ini[t] + INFIL[t]
    ET_soil[t] = S * (1 - exp(- ( ET_pot[t] - ET_dir[t] ) / K ))
    S          = S - ET_soil[t]
    PERC[t]    = k * S
    S          = S - PERC[t]
    Q_soil[t]  = max( S - K , 0 )
    S_fin[t]   = S - Q_soil[t]

    ### SOIL SATURATED ZONE PROCESSES ###
    G         = G_ini[t] + PERC[t]
    Q_base[t] = lamda * max( G - H , 0 )
    G         = G - Q_base[t]
    Loss[t]   = a * G
    G_fin[t]  = G - Loss[t]

    if ( t == t_max ) { break }
```

```
  else  {
    S_ini[t+1] = S_fin[t]
    G_ini[t+1] = G_fin[t]
  }

}

ET_act = ET_dir + ET_soil
Q_sim  =  Q_dir +  Q_soil + Q_base


#### NASH - SUTCLIFFE EFFICIENCY (NSE) ####

### CALIBRATION PERIOD ###
for ( i in 1:Cal_Per ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[1:Cal_Per] ) )^2
}
NSE_Cal = 1 - sum( dQ[1:Cal_Per] ) / sum( dQ_obs[1:Cal_Per] )

### VALIDATION PERIOD ###
j = Cal_Per + 1
for ( i in j:length(P) ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[j:length(P)] ) )^2
}
NSE_Val = 1 - sum( dQ[j:length(P)] ) / sum( dQ_obs[j:length(P)] )

### WHOLE PERIOD ###
for ( i in 1:length(P) ) {
  dQ[i]     = ( Q_obs[i] - Q_sim[i] )^2
  dQ_obs[i] = ( Q_obs[i] - mean( Q_obs[1:length(P)] ) )^2
}
NSE_Tot = 1 - sum( dQ[1:length(P)] ) / sum( dQ_obs[1:length(P)])


#### MODEL OUTPUT ####

column.names = c('NSE_Calibration','NSE_Validation','NSE_Total')

resultsTable = array(c(format(NSE_Cal, digits = 4, width = 11,
                              justify = 'centre'),
                       format(NSE_Val, digits = 4, width = 10,
                              justify = 'centre'),
```

```
                        format(NSE_Tot, digits = 4, width = 7,
                               justify = 'centre')),
                      dim = c(1,3), dimnames = list('Value',column.names))


  row.names    = seq(from = 1, to = t_max, by = 1)
  column.names = c('Simulated_Runoff','Actual_ET',
                   'Soil_Moisture','Groundwater_Storage')

  operationTS  = array(c(format(Q_sim,  digits = 3, width = 10,
                                justify = 'centre'),
                         format(ET_act, digits = 3, width = 7,
                                justify = 'centre'),
                         format(S_fin,  digits = 3, width = 11,
                                justify = 'centre'),
                         format(G_fin,  digits = 3, width = 14,
                                justify = 'centre')),
                       dim = c(t_max,4), dimnames = list(row.names,column.names))

  class(operationTS) = 'numeric'

  output = list(resultsTable , operationTS)

  print(resultsTable, quote = FALSE)

  return( output )

}
```

Appendix B: R script for Monthly Time Scale Analysis with "Zygos6P" model

Appendix B1*: Main Script*

```
#############################################################################
#                                MAIN SCRIPT                                #
#############################################################################

setwd('C:/Users/User/Documents/R/HydroR/MSc_R_Code')

historical_data = read.csv('C:/Users/User/Documents/R/HydroR/MSc_R_Code/
                           Loing_monthly_data.csv', header = TRUE)
synthetic_data = read.csv('C:/Users/User/Documents/R/HydroR/MSc_R_Code/
                           Loing_month_stoch_sim2.csv', header = TRUE)
```

```
message('\n')
C = as.character( readline(prompt = 'Please enter desired calibration scheme ("SS"
                                    for Split-Sample & "SC" for Stochastic
                                    Calibration) : \n\n') )


if (C == 'SS') {

  Date   = as.Date(historical_data$Date, format = '%m/%d/%Y')

  P      = historical_data$P          # Precipitation
  ET_pot = historical_data$PET        # Potential Evapotranspiration
  Q_obs  = historical_data$Qobs       # Observed Runoff

  message('\n')
  p = as.integer( readline(prompt = 'Please enter desired calibration sample as a
                                    percentage (%) of the whole length of data
                                    records : \n\n') )

  Cal_Per = as.integer( round( (p/100) * length(P) ) )      # Calibration Period
                                                            # [in months]

  message('\n')
  message(sprintf('Calibration Period : %.1f percent of the whole length of
                  data records.', p))
  message('\n')

} else {

  Date   = as.Date(synthetic_data$Index, format = '%m/%d/%Y')

  P      = synthetic_data$Rainfall    # Precipitation
  ET_pot = synthetic_data$PET         # Potential Evapotranspiration
  Q_obs  = synthetic_data$Flow        # Observed Runoff

  Cal_Per = as.integer( length(P) )   # Calibration Period [in months]

  message('\n')
  message('Calibration Period : whole length of synthetic data')
  message('\n')

}

#### MODEL CALIBRATION USING OPTIMIZATION ALGORITHM 'EAS' ####
```

```
n = 6     # Number of Model Parameters
m = 100   # Population Count

xmin = c( 0.1,  100,   0,    0,   0, 0.01 ) # Inner Lower Boundaries of
                                            # Model Parameters
xmax = c(   2,  300, 0.3,  100, 0.2, 0.30 ) # Inner Upper Boundaries of
                                            # Model Parameters
xlow = c( 0.1,  100,   0,    0,   0, 0.01 ) # Outer Lower Boundaries of
                                            # Model Parameters
xup  = c( 2.5, 1000, 1.0,  300, 1.0, 1.00 ) # Outer Upper Boundaries of
                                            # Model Parameters
maxeval   = 1000
ratio     = 0.99
pmut      = 0.5
beta      = 2
maxclimbs = 5
ftol      = 0

x = eas(n, m,
        xmin, xmax, xlow, xup,
        fn = ObjFun,
        maxeval, ftol, ratio, pmut, beta, maxclimbs,
        P = P, ET_pot = ET_pot, Q_obs = Q_obs, Cal_Per = Cal_Per,
        Change_Sing = TRUE)

sol  = x[['bestpar']]
PARS = c(sol[1],sol[2],sol[3],sol[4],sol[5],sol[6])


#### MODEL EFFICIENCY & TIME SERIES ####

if (C == 'SC') {      # SC: Validation against the whole historical sample

  Date   = as.Date(historical_data$Date, format = '%m/%d/%Y')

  P      = historical_data$P          # Precipitation
  ET_pot = historical_data$PET        # Potential Evapotranspiration
  Q_obs  = historical_data$Qobs       # Observed Runoff

  Cal_Per = as.integer( round( (p/100) * length(P) ) )      # Calibration Period
                                                            # [in months]

}
```

```
output = Zygos6P(PARS , P , ET_pot , Q_obs , Cal_Per)

# output      : List of 2 tables
# output[[1]] : NSE Results for Calibration, Validation and Total Data Records
# output[[2]] : Time Series for Simulated Runoff, Actual ET, Soil Moisture &
#               Groundwater Storage

resultsTable = output[[1]]
operationTS  = output[[2]]

message('\n')
message('~ Optimal values for Model Parameters ~ \n\n')
message('------- Model Parameters ------- \n')
message(sprintf(' v             =          %5.3f \n', PARS[1]))
message(sprintf(' K             =       %5.1f mm \n', PARS[2]))
message(sprintf(' k             =          %5.3f \n', PARS[3]))
message(sprintf(' H             =       %5.1f mm \n', PARS[4]))
message(sprintf(' a             =          %5.3f \n', PARS[5]))
message(sprintf(' lamda         =          %5.3f \n', PARS[6]))
message('------------------------------- \n\n')

print(resultsTable, quote = FALSE)
```

Appendix B2: *Objective Function*

```
##############################################################################
#                       FUNCTION : Objective Function                        #
##############################################################################

ObjFun = function(Pars , P , ET_pot , Q_obs , Cal_Per, Change_Sing)  {

  output = Zygos6P(Pars , P , ET_pot , Q_obs , Cal_Per)

  resultsTable = output[[1]]

  NSE_Cal = as.numeric(resultsTable[1])

  if ( Change_Sing == 1 ) { NSE_Cal = - NSE_Cal }

  return(NSE_Cal)

}
```